University of Pisa
Research Doctorate School in Biological
and Molecular Sciences

Doctorarate Program: Biology

The DNA barcoding in the ciliated protist *Euplotes*

Supervisor:
Dr. Ph.D. Graziano Di Giuseppe

Student:
Dr. Paola Ricciolini

Academic Year 2011/2012
BIO/05

1 ABSTRACT 4

2 INTRODUCTION 6

2.1 THE TAXONOMIC IMPEDIMENT 7

2.2 TOWARDS A SOLUTION: THE DNA BARCODING TOOL AND ITS POTENTIAL 9

2.2.1 THE DNA BARCODING TOOL AND ITS BROAD RANGE OF APPLICATIONS 12

2.2.2 THE UNIVERSAL DNA BARCODE: THE *cox1* GENE 14

2.2.3 THE TWO CRITERIA FOR ASSIGNING A DNA SEQUENCE TO THE SPECIES LEVEL 16

2.2.4 THE DNA BARCODING: "PHILOSOPHICAL" AND TECHNICAL CONSIDERATIONS 19

2.2.5 THE DNA BARCODING: STATE OF THE ART 23

2.3 PROTISTS: GENERAL FEATURES 26

2.3.1 PROTISTS AND THEIR TANGLED TAXONOMY 28

2.4 CILIATED PROTIST: GENERAL FEATURES 30

2.4.1 THE CILIATED PROTIST GENUS *Euplotes* AND ITS TAXONOMIC IMPEDIMENT 31

3 AIM OF THIS WORK 34

4 MATERIALS AND METHODS 36

4.1 SOURCE OF SAMPLES AND CULTURE CONDITIONS 37

4.2 DNA EXTRACTION 44

4.3 PRIMER DEVELOPMENT, PCR AMPLIFICATION, AND SEQUENCING 46

4.4 MULTIPLE SEQUENCE ALIGNMENT (ALN) 54

4.5 TESTING FOR MUTATIONAL SATURATION 55

4.6 PHYLOGENETIC ANALYSIS 56

5 RESULTS 58

5.1 SEQUENCES OBTAINED 59

5.2 NUCLEOTIDE COMPOSITION 63

5.3 MULTIPLE ALIGNMENT (ALN) 65

5.4 SEQUENCE ANALYSIS OF THE *cox1* GENE 68

5.5 TESTING FOR MUTATIONAL SATURATION 74

5.5.1 LIKELIHOOD-MAPPING ANALYSIS 74

5.5.2 p-distance ANALYSIS 76

5.6 PHYLOGENETIC ANALYSIS 78

6 DISCUSSION 90

6.1 SEQUENCES ANALYSIS OF THE *cox1* GENE 91

6.2 PHYLOGENETIC ANALYSIS 93

7 CONCLUSIONS 97

8 LITERATURE CITED 99

Questa Tesi di Dottorato

la dedico con tanta ammirazione

alla Dottoressa Silvia Frosini

e alle sue determinazione e forza di volontà.

Io al tuo posto non ce l'avrei fatta.

1 ABSTRACT

Recent investigations have suggested the feasibility of establishing a species identification system reliant on the analysis of the sequence of the mitochondrial cytochrome *c* oxidase subunit 1 gene (*cox1*), the DNA barcoding tool. I assessed the effectiveness of this procedure in the most cosmopolitan, ubiquitous, and differentiated group of ciliated protists, the genus *Euplotes*. The availability of the largest collection of *Euplotes* living strains which currently exists worldwide, the result of half a century of sampling at all latitudes of the globe, allowed me to help create the basis for making this evaluation. Accordingly, I analyzed *cox1* gene sequences from 81 strains belonging to 15 *Euplotes* species representatives of different habitats. All *Euplotes* strains analyzed were correctly partitioned into the 15 species, whose interspecific *cox1* sequence divergence was about 60%, whereas the intraspecific sequence divergence ranged from 0.5% to 43%. Moreover, some species not only showed higher values of *cox1* intraspecific divergence, but also their representative strains did not cluster together on the *cox1* phylogenetic trees, suggesting the occurrence of cryptic species. Overall, my study demonstrates the feasibility of the DNA barcoding procedure to the genus *Euplotes*.

2 INTRODUCTION

2.1 THE TAXONOMIC IMPEDIMENT

Biodiversity means the variety of life forms that inhabit the planet Earth. Biodiversity is our natural and, unfortunately, vulnerable resource. It is in our own interest to preserve this resource, in order to benefit from it. To be able to protect it, we must first know it. In fact, among the most urgent needs and ambitions of our society, there is to know in detail and catalog all the species of the planet Earth. This is necessary to various reasons. The first reason is because of our increasing interaction with the biodiversity that surrounds us. The most important reason is because we are facing the greatest loss of this biodiversity that has ever occurred. For this final reason the year 2010 was proclaimed by the United Nations (UN) Year of Biodiversity. We need to protect endagered species, to ensure bio-security avoiding pandemics, and to manage the natural world. The first premise of all biological researches is the correct species identification. In particular, over the past 15 years, the need to study more thoroughly the variety of life forms of microorganisms has been felt. This is due because the biological diversity of microorganisms represents the majority of biodiversity on our planet. Until a few years ago, it was not possible to study the biodiversity of microorganisms, because of the lack of methodologies to do so. The advent of modern molecular biology techniques, especially the Polymerase Chain Reaction (PCR), has overcome this obstacle. In recent years, in fact, numerous studies have been conducted on the biological diversity of microorganisms. These studies have increased the interest in biodiversity of eukaryotic microorganisms. The interest of the scientific community in the biological diversity of eukaryotic microorganisms is due to various reasons. The first reason is that they play a leading role in the biogeochemical cycles of the Earth. The second reason is that they are utilized in the research for new bioactive molecules for different purposes [1-3].

To complete this inventory of life is a gargantuan enterprise, because of the huge number of species on the Earth. To accomplish this purpose, our society needs an accurate diagnostic tool which is faster and cheaper to study our biodiversity on a large scale, in all the present environments, and possibly retrieve some information about the past and make prediction about the future. Unfortunately, we still lack this tool. It is clear that Taxonomy and Systematic, the biological disciplines that are devoted to characterizing the diversity of life and organizing our knowledge about this diversity, play a practical role in directing the preservation and development of natural systems [4]. Carl Nilsson Linnaeus (1707-1778), known to most simply as Linneo, is

considered the father of modern scientific classification of living organisms thanks to the introduction of binomial nomenclature. Unfortunately, during the last 250 years, we have described only a small fraction, about 1.7 million species, of the estimated biodiversity, 10-100 million species [5, 6].

The morphological species concept is the most broadly used in eukaryotic organisms. Consequently, the actual identification system of species is based primarily on the morphological approach based on morphological features or "keys": shape, size, colour of body parts, etc… It has served us well for a little over three centuries, but is characterized by severe limitations, making it inadequate for current needs. Therefore we speak of "taxonomic impediment" (i.e., the worldwide shortage of taxonomists and their declining rate of replacement) that inhibits biodiversity research [7, 8]. Almost no taxonomic research projects are still funded. Moreover, scientific journals are becoming less interested in publishing articles in this field. Hebert and collaborators (2003) [9] discussed the limitations inherent to the morphology-based taxonomic approach summarizing them in four main points. (I) This method is particularly difficult, requires expert skills and proceeds too slowly. (II) Morphological keys that enable identification are not always present in all file stages or gender and they are not always sufficient to determine species boundaries. For example, in Diptera the species identification is mainly based on male genitalia. (III) Phenotypic plasticity greatly complicates the analyses. (IV) This approach also overlooks the morphologically cryptic taxa, which are common in many groups. Taken all together these features can lead to incorrect identifications and so misdiagnoses are common. Furthermore, if nothing is done to change the slow pace of current taxonomic efforts and practice, it will take centuries to complete even a preliminary "Encyclopaedia of life" on Earth [10]. For example, regarding unicellular eukaryotic microorganisms (protists), the relationship between morphology and species boundaries is not well understood. Furthermore, many protistan species are morphologically indistinguishable. Consequently, we speak of "cryptic diversity" (i.e. biodiversity that is not reflected in morphological features).

2.2 TOWARDS A SOLUTION: THE DNA BARCODING TOOL AND ITS POTENTIAL

Tautz and collaborators (2003) [11] explored the possibility of overcoming the problems reliant on the morphology-based identification system by creating an identification system for all living organisms based on short DNA sequences, a so "DNA-based taxonomy" system. This is not new in essence; in fact there is a long history of using molecular markers: allozymes, and both nuclear and mitochondrial genes for species identification. Ribosomal DNA (rDNA) coding for ribosomal RNA (rRNA) is present in all eukaryotic cells. For this reason, nuclear ribosomal genes such as the 18S ribosomal RNA (18S rRNA) gene and ITS (Internal Transcribed Spacer) region are broadly used in phylogeny. Especially as regards the 18S gene, there is a wide literature of inferring phylogenetic relationships from this gene, mostly in animals, but also in unicellular eukaryotic microorganisms, the protist. It owes its success as a phylogenetic marker to the fact that it codes for a basic component of all eukaryotic cells, and its flanking regions allow the designing of universal PCR primers. However, after initial enthusiasm, there were also some limitations inherent in the use of the 18S gene. Mitochondrial genes coding for rRNA 12S and 16S were used too, but the prevalence of insertions and deletions greatly complicate sequence alignments. So there is nothing fundamentally new in the "DNA-based taxonomy" concept. However, all current taxonomic approaches intend to use DNA, at best, as an auxiliary criterion for identifying a species or a taxon, but have not given it a central role. The proposal of Tautz and collaborators (2003) [11] is innovative because they thought of a DNA-based identification system which is "the central pillar of taxonomy, whilst maintaining the importance of morphological criterion".

"DNA barcodes" is an old term found for the first time in literature in 1993 [12], in a paper that has gone unnoticed. Hebert and collaborators rediscovered the DNA barcoding concepts in a paper published in 2003 [9]. The DNA barcoding is a molecular method that uses a short genetic marker (i.e. DNA barcode) in an organism's DNA to identify it as belonging to a particular species. In a very real sense, these sequences can be viewed as "genetic barcodes" that are embedded in every cells and were similar to the way a supermarket scanner distinguishes products using the black stripes of the Universal Product Code. The underlying assumption is that the genetic variation between species exceeds that within species. The Authors clearly highlighted the power of this approach to species identification when phenotypic plasticity is a concern, morphology keys are not

available or unknown, and cryptic species. The other main purpose of this approach is to discover new species [13].

What makes the DNA barcoding technique conceived by Hebert and collaborators [9, 13] an innovative tool is the idea to radically replace classical taxonomy with this new approach, and the large scale of its technological and social ambitions. Technological, in the sense that the method uses the same standard and inexpensive experimental pipeline, the same DNA barcode, and the same criteria to mark species boundaries **(Fig. 1)**. Social, in the sense that the DNA barcoding tool is born to be easily used by all, not only by experts, and in the desire to make the data accessible to all. In fact, the DNA barcoding is not only a new tool to identify species, but it is an ambitious international project too: the international Barcode Of Life project (iBOL, http://www.ibol.org), the largest biodiversity genomics initiative ever undertaken. The Barcode of Life Data Systems (BOLD) is a computerized workbench for the acquisition, storage, analysis and publication of the DNA barcoding records. This database was created and is maintained by the University of Guelph in Ontario. This database is public and freely available to anyone interested in DNA barcoding. Via the World Wide Web, scientists and the general public have direct access to all life form information, not only the barcode sequences: images, biology, and conservation status **(Fig. 1)**. This allows to greatly speed up communication, making species diagnoses and new descriptions more accessible [14, 15]. The Consortium for the Barcode Of Life (CBOL, http://barcoding.si.edu) is an international project aimed to promote and coordinate the DNA barcoding research in over fifty countries by promoting workshops, conferences, training and more. The CBOL was established in 2004 and its main proposal is the rapid and inexpensive identification of the estimated 10 million species on Earth.

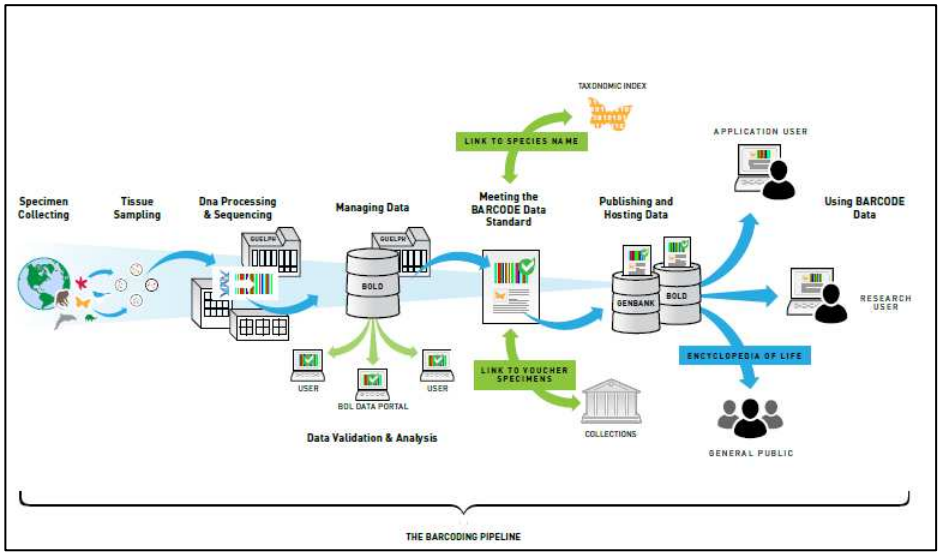


Fig. 1. The standard DNA barcoding pipeline has four components (http://www.ibol.org). **(I)** The first step is the collection of the specimen belonging to an unknown species. Specimens may have come from various sites: sampling, museums, collections, herbaria, zoos, aquaria, and other repositories of biological materials. **(II)** The second step is the laboratory analysis, in order to obtain the DNA barcode sequence. **(III)** Analysis of the barcode sequence allows the specimen species identification. **(IV)** The last step is the sharing of the data by submitting them in the on-line BOLD database.

2.2.1 THE DNA BARCODING TOOL AND ITS BROAD RANGE OF APPLICATIONS

The DNA barcoding tool has a wide range of applications, not only in Taxonomy and Systematic.

Ecologists in particular are great supporters of the DNA barcoding tool [16, 17]. Ecologists will be able to not only identify a single species from a specimen or an organism's remains, but also determine the species composition of environmental samples. Several taxa have now been surveyed in their natural habitats using this technique; short DNA fragments persist in the environment and might allow an assessment of local biodiversity from soil or water [18-22]. This is especially important to study the biodiversity of microorganisms. In fact, until now the major limitation in their studying was the fact that not all microorganisms can be cultivated in the laboratory. In the rainforests, rapid DNA-based entomological inventories have been efficiently performed [16, 23-25]. In forensic sciences, the DNA barcoding can help in solving crime investigations by determining the time of death by inferring the life stage of the insect larvae within the dead body [26]. It can be used for understanding interspecies interactions, for example the DNA barcoding approach has already shown that the existence of cryptic species could mask the specialization of a parasite to a single host [27, 28].

Furthermore, the DNA barcoding can be advantageous for monitoring illegal trade in animal products: food, luxury accessories, and clothes [29-31]. In Africa, many species of mammals, especially primates, are threatened by illegal trade in bush meat. The DNA barcoding in such cases could be usefully employed to rapidly identify to which organism the minced or smoked meat belongs. It would not be identifiable in any other way. The DNA barcoding can provide useful information for governments about population sizes in order to manage and monitor natural ecosystems. Such an approach is now widely used, and is particularly useful for detecting the presence of elusive or endangered species [17]. For example, it is already being used to combat illegal hunting of cetaceans [32].

Moreover, DNA barcoding tool can be used in the field of biosecurity. For example, it can be used in the identification and surveillance of exotic species and for surveillance of disease vectors, such as invasive insects [33, 34]. Particularly, in insects, a pest at the egg or larval stage

might not be recognizable without DNA identification [22]. Every year millions of dollars are thrown away because of agricultural pests. It can also be usefully employed to monitor water quality. In addition, ecologists can take advantage of DNA tools when only hair, feces or urine left behind by animals are available for species identification. Even DNA-based diet composition can be estimated using faecal samples [17].

The DNA barcoding can also enhances discovery of new species [13]. Using high-throughput methods developed for genomic studies, potentially hundreds of new species could be discovered weekly from environmental samples, especially from the depths of the oceans.

Concluding, the DNA barcoding is potentially a simple, fast and effective investigative tool, and accordingly appears to be particularly interesting for those who work with the least morphologically tractable groups, that are protists [35, 36].

2.2.2 THE UNIVERSAL DNA BARCODE: THE *cox1* GENE

In the previous paragraph I discussed the potential of the DNA barcoding tool as a universal standard method. This implies the use of the same "DNA barcode" for the identification of all animal species on the Earth. The DNA barcode is a unique segment of DNA that can identify an organism, or part of an organism. Hebert and collaborators (2003) [9] first established the utility of a ~ 650 base pair (bp) fragment of the mitochondrial cytochrome *c* oxidase subunit 1 (*cox1*) gene as universal marker (or DNA barcode) for global biological identification of animal species **(Fig. 2)**.

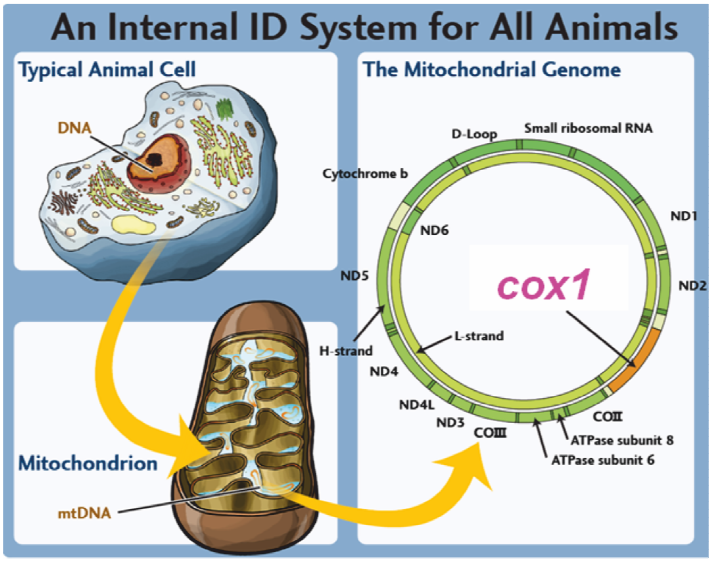


Fig. 2. The *cox1* gene as "gold standard" for global biological identification of animal species.

The DNA barcode itself consists of a 648 bp region comprised between 58–705 bp from the 5'-end of the *cox1* gene using the mouse mitochondrial genome as a reference.

The steps that led to this decision are described below. Mitochondria are characteristic organelles of eukaryotic cells. They are deputies to the production of energy in the form of ATP. They have their own genome (mtDNA), which is contained in multiple copies within the mitochondria. The mtDNA has been used extensively in the last years for inferring the genetic structure of a population. Furthermore, in recent years it has also been used for inferring phylogenetic studies. The pilot researches about mtDNA variation focused on the restriction

fragment length polymorphisms (RFLP) [37-39]. These researches marked the advent of the mtDNA as a molecular evolution study tool. The general features of mitochondria have been studied deeper in higher metazoa. However, animal mtDNA is now considered more variable in size and gene organization than was previously thought. Normally, it is a circular molecule, containing both coding and transcribed genes in various kinds of RNA, such as tRNA and rRNA. There are also exceptions. For example, in the ciliated protists *Euplotes crassus* and *Euplotes minuta,* and in some algae, their mitochondrial genomes are linear [40]. From the beginning, the research for the universal barcode has focused on the mitochondrial genome, because as compared to the nuclear one (nuDNA), it has features that have attracted researchers. (I) Higher rate of DNA evolution because of frequent replications, and absence of both histones and repairing mechanisms. (II) Limited exposure to recombination due to its maternal mode of inheritance. That is important because most methods of phylogenetic reconstruction assume no recombination. (III) Lack of introns, which greatly complicate sequence alignments across phyla. (IV) As it is present in higher copies number per cells allows for greater efficiency of DNA extraction [41].

Hebert and collaborators (2003) [9] justify the choice of a mitochondrial protein-coding gene because its third position nucleotides show a high incidence of base substitutions, leading to a rate of molecular evolution that is about three times greater than that of other mitochondrial non-coding genes, such as the 12S or the 16S rRNA genes [42]. Moreover, the utility of these genes in the phylogeny is curtailed by the presence of insertions and deletions, which greatly complicate sequence alignments [43]. There are 13 protein-coding genes in the animal mitochondrion genome, but Hebert and collaborators (2003) [9] point to two advantages of the *cox1* gene: (I) the universality of existing primers for amplification of the 5’ end of this gene in a wide variety of animals [44, 45], and (II) the broad phylogenetic range covered by this gene. In fact, these Authors argued that the evolution of the *cox1* gene is rapid enough to allow the discrimination of not only closely allied species, but also phylogeographic groups within a single species [46, 47]. Although the *cox1* gene may be matched by other mitochondrial genes in resolving such cases of recent divergence, this gene is more likely to provide deeper phylogenetic insights than alternatives such as cytochrome b gene [48].

2.2.3 THE TWO CRITERIA FOR ASSIGNING A DNA SEQUENCE TO THE SPECIES LEVEL

The first criterion for assigning a DNA sequence to the species level is based on the monophyletic species concept. Sequences for the same species are considered to be correctly identified as long as they form a monophyletic cluster on a phylogenetic tree **(Fig. 3)**. The DNA barcoding research only used the *cox1* gene and analyzed data by only Neighbor-Joining (NJ) [49] phylogenetic inference, using the Kimura 2 Parameter (K2P) distance model of sequence evolution [50].

The second criterion is based on a threshold **(Fig. 4)**. This threshold is set to differentiate between intraspecific and interspecific *cox1* nucleotide sequence variation, the "DNA barcoding gap". Because patterns of intraspecific and interspecific *cox1* sequence variation appear similar in various animal groups, Hebert and collaborators (2004) [13] proposed a universal threshold to mark the boundaries between species: ten times the mean intraspecific variation for the group under study. If two sequences differ from each other by a value under ten times the mean intraspecific *cox1* sequence variation, they are recognized as belonging to the same species. Otherwise, they are recognized as belonging to two different species.

The first papers published on DNA barcoding followed these guidelines for data analysis. However, the articles published later do not necessarily use the algorithm NJ. In fact, the Maximum Likelihood (ML) method and the Bayesian Inference (BI) are becoming increasingly used. Furthermore, not all researchers applied the threshold method. This is mainly due to the lack of possibility to establish a well-defined threshold to mark the boundaries between species.

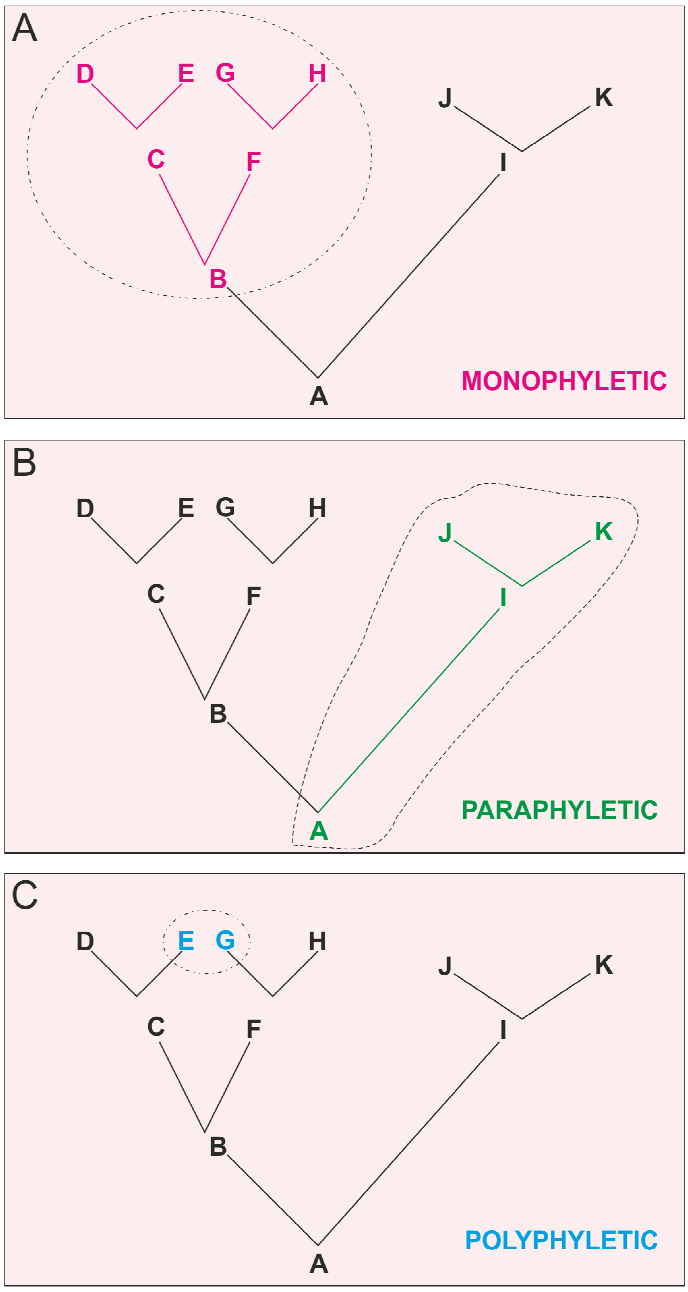


Fig. 3. (A) Monophyletic group is a taxon which forms a clade on a phylogenetic tree. A clade is a group consisting of a species and all its descendant forming a single branch in phylogenetic trees. **(B)** Paraphyletic group includes the most recent common ancestor, but not all of its descendents. **(C)** Polyphyletic taxon does not include the common ancestor.

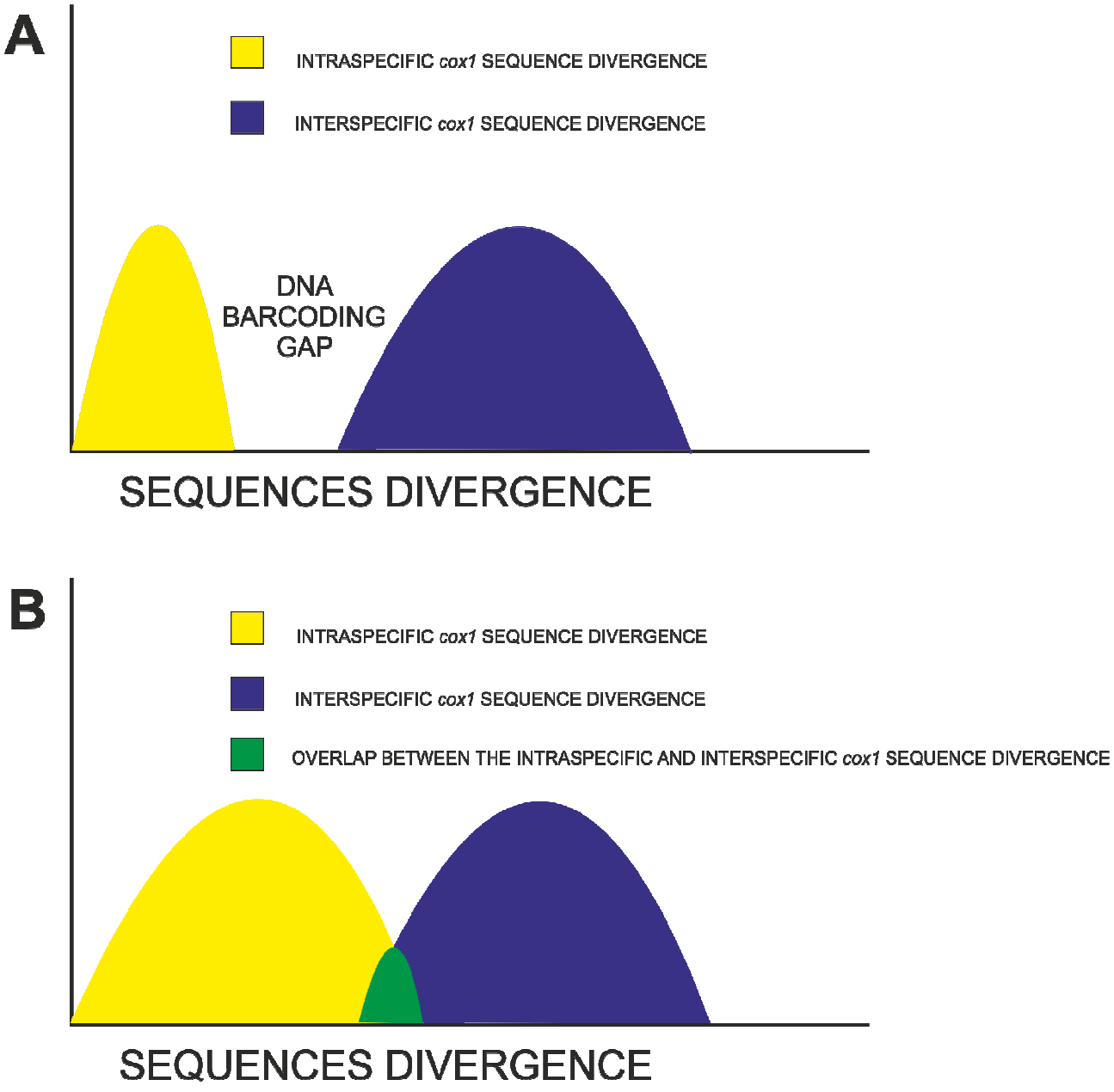


Fig. 4. (A) Intraspecific and interspecific *cox1* gene nucleotide sequence divergence values do not overlap. Consequently, the species can be distinguished on the basis of the threshold method. **(B)** Intraspecific and interspecific *cox1* gene nucleotide sequence divergence values overlap. Consequently, the species cannot be distinguished on the basis of the threshold method.

2.2.4 THE DNA BARCODING: "PHILOSOPHICAL" AND TECHNICAL CONSIDERATIONS

The use of the DNA barcoding technique remains among the most contentious and many Authors have clearly expressed their concerns. Concerns range from the "philosophical" to the technical. Only mentioning the possibility to radically replace the traditional morphology-based taxonomic approach with the DNA-based taxonomy based on a single mitochondrial gene and on a universal threshold has generated lively debates. The taxonomists feel pushed aside and insulted by this new technology and have renamed it with colorful nicknames: the new Tower of Babel, a serious moving backwards for Science, a caricature of the real Taxonomy, and a sterile intellectual landscape [51-56]. These Authors discussed the fact that often, in literature the DNA barcoding didn't work well and/or revealed unexpected diversity or discordance with morphology. It is true that the DNA barcoding is limited to matching DNA sequences to known species, the latter being delimited with traditional (e.g. morphological) methodologies. The role of barcodes is merely to provide a tool to assign unidentified specimens to already characterized species. The DNA barcoding is based on knowledge of Taxonomy, and cannot function without that knowledge: if species are not well defined and thoroughly sampled, the DNA barcoding is not applicable. For those groups where barcoding would be useful, the vast majority of the taxonomic work is already done, so the role of this tool is merely to give a confirmation of taxonomic work. This is true, but it helps in finding cryptic species, which are not detected by the classical approach. These Authors argued that the DNA barcoding will not help with the ''taxonomic impediment'', that can only be solved through an understanding of complex species definitions and effective and accurate biodiversity assessments. Moreover, they highlighted that in some cases the DNA barcoding also fails to distinguish between members of closely related species groups and morphologically highly similar species [57], two instances for which the DNA identifications should have had the most value. The following is an impartial discussion of the reasons why many Authors consider mtDNA divergence neither necessary nor sufficient as a criterion for delineating species boundaries.

The very first point of objection to this methodology is the assumption that one single mitochondrial gene sequence should be the primary identifier for species. In fact, many Authors agree that it must be considered in conjunction with other sources of data such as nuclears genes, morphology, and ecology [58]. First, not all organisms are equipped with mitochondria, and

therefore in these cases this practice cannot be applied. Rates of evolution for the same gene in different taxonomic groups may vary, making the use of any single gene problematic as "the gold standard" for molecular taxonomy across very broad phylogenetic boundaries. A combination of genes may prove useful for supporting a purely molecular taxonomy. Concerns have been expressed regarding horizontal transfers of mitochondria between divergent lineages [53]. Those transfers can occur between closely related species too, and could result in wrong diagnoses [13, 59, 60]. An additional problem with focusing on a single DNA sequence as a primary criterion for recognizing species, is that it will lead us to overlook new or rapidly diverged species [61]. For some time after the initial split, new sister species will share alleles, either because of ongoing gene flow, or because of recent ancestry. In such cases, sequences from one or few individuals will not be sufficient for an unequivocal assignment to a particular group. Yet, it is not true that mitochondrial recombination does not occur; in fact recent investigations have found significant evidence of mitochondrial recombination in various animal groups [62-64]. The recombination can affect the NJ analysis. Moreover, in mitochondria the mutation rate of a gene is a function of its physical location in the genome [41, 65]. Thus, variations in genome arrangement will affect the mutation rate of the *cox1* gene, and so, the analysis. We also must consider that the substitution patterns are not constant along the length of the gene [66, 67]. Moreover, significant levels of heteroplasmy (carrying more than one mtDNA haplotype) have been reported in various animal groups [68-74]. One single different nucleotide could alter the results. Nuclear pseudogenes could be amplified by PCR instead of the real mitochondrial gene [75]. However, BOLD provides quality controls in order to reveal the presence of pseudogenes within the dataset. In addition, criticism also applies to the choice of the *cox1* gene as universal barcode, because of its limited information content at deeper phylogenetic levels [76, 77]. Only two studies deal with the problem of the choice of the region of the *cox1* gene to be analyzed [67, 78]. In particular, Roe and Sperling (2007) [67] emphasized that the choice of amplifying the ~ 650 bp *cox1* gene fragment to the 5' end, was merely based on the availability of the metazoan universal PCR primers of Folmer [44]. This theoretical decision is lacking any biological issue, and in the end no one used these primers, because in reality they are not universal. Consequently, researchers had to design taxon-specific primers. Accordingly, Roe and Sperling (2007) suggested to extending the analysis to a longer portion of the gene. In fact, these Authors, despite the claim of Hebert and Gregory [79] according to which "DNA barcoding is not intended to reconstruct phylogenetic relationships, but instead focus explicitly on species delimitation and diagnostics", are worried about heterogeneous substitutions patterns along the gene, and because the

problem of DNA saturation. The DNA saturation is the lost of DNA information for phylogenetic purpose, due to multiple hints occurring in the same positions. Roe and Sperling (2007) concluded that the best choice, in order to reflect broader patterns of nucleotide divergence and minimize nucleotide saturation, is maximizing the analyzed sequence length. Regarding this issue, Erpenbeck and collaborators (2005) [78] demonstrated that using the 3' portion of the *cox1* gene as "gold fragment" for DNA barcoding in corals, allows species identification. Otherwise its resolution is limited to genus level. Therefore, many Authors agree that a multi-markers approach, based on both nuclear and mitochondrial genes, is the best choice in order to infer the best phylogeny. Bittner and collaborators (2010) [80] argued that such multi-markers DNA barcoding is the same of a metagenomic approach, a powerful tool to study biodiversity on large scale and on environments. The choice about the number and identity of genes necessary for sufficient resolution may be different for different clades. Furthermore, also Rubinoff and collaborators (2006) [81] stated that "good phylogenetic are absolutely essentials".

Another line of criticism concerns how the data are analyzed. Here, again, is the problem of inferring phylogenetic trees from one single gene. It is well documented, in fact, that multiple dataset allow clearest definition of phylogenetic relationships among taxa [82-84]. Hebert and collaborators (2003) [9] proposed the use of the K2P distance model of sequence evolution and NJ phylogenetic inference. Srivathsan and Meier (2011) [85] defined as inappropriate the use of the K2P model in the DNA barcoding literature. The NJ analysis is the simplest and faster one, but reliable estimates of pairwise distances can be hard to obtain for divergent sequences [86]. Therefore, some Authors are suspicious of this type of phylogenetic reconstructions [54], suggesting other methods for inferring phylogenetic trees, such as Maximum Parsimony (MP), Bayesian Inference (BI), and Maximum Likelihood (ML) [87]. However, other Authors did not show any difference in accuracy between the NJ, MP and ML analysis of DNA barcoding of tropical butterflies [88]. Some Authors also highlighted that the threshold method lacks any biological and especially taxonomic justification. Moreover, it did not work well because there was a broad overlap of interspecific and intraspecific distances [81]. The accuracy of the DNA barcoding tool in identifying species depends on the breadth of the DNA barcoding gap, and the more overlap there is between the interspecific and the intraspecific genetic variation, the less effective the tool becomes, and then there is the need to establish a case-specific threshold [87]. Some Authors are also worried about the under sampling, both in numerical terms and in terms of

too geographically restricted sampling [89]. The accuracy of the threshold methods depends upon sampling multiple specimens from across the known geographic range of the species.

2.2.5 THE DNA BARCODING: STATE OF THE ART

Despite all the skepticism, so far the *cox1* gene has been proved to be suitable for the identification of a large range of animal taxa, and the number of the DNA barcoding projects is growing up. Moreover, it has revealed the presence of a large cryptic or pseudocryptic diversity [13, 90, 91]. It follows that the *cox1* gene seems to be the ideal candidate for use as the universal marker or barcode in the animal organisms, with the exception of the Cnidaria phylum. This is because the rates of mitochondrial evolution are exceptionally low in these organisms, and therefore they don't enable species discrimination [92, 93]. The DNA barcoding worked well especially in higher metazoa: ants [24], birds [13, 94], bryozoans [95], butterflies [9, 57], crustacea [96], fish [21], gastropods [97], millipedes [98], nematodes [99], primates [100], and spiders [101]. In bryozoans, the DNA barcoding technique performed great in discriminate species. Moreover, it revealed a high degree of cryptic diversity [95]. Also Spelda and collaborators (2011) [98] explained the high intraspecific *cox1* variation in some millipedes genera with the occurrence of cryptic lineages. Nijman and Aliabadian (2008) stated that the use of mitochondrial markers in studying humans and primates evolutionary biology is gaining importance. They also demonstrated that the *cox1* gene allowed a clearer separation between species than the mitochondrial 16S and cytochrome b genes [100].

However, it is very important to point out that the *cox1* gene is the selected barcode for the animal organisms. Outside those ones, it has not yet exhaustively tested if the *cox1* gene will function for species discrimination. In the other living organisms the mode of inheritance, rate of divergence, and both mitochondrial genome and *cox1* gene are poorly known. In any case, regarding land plants, it was well established that mitochondrial genes are generally more slowly evolving than those in animals [102]. The *cox1* gene was useful only in some macroalgae [103-105]. Consequently, researchers explored other barcode in land plants. Some Authors proposed the use of a multi-markers approach [106, 107]. Plastid markers such as the megakaryocyte-associated tyrosine kinase (*matk*) and the ribulose bisophosphate carboxylase (*rbcL*) genes, were proposed as genetic barcodes for this multi-markers approach in land plants [108]. This approach has also received considerable interest from mycologists [109]. In fungi, there was a big debate about who is the ideal barcode. The *cox1* gene has been proved to be suitable for the identification of fungi, suggesting that the feasibility to extend this approach to fungi is high [110]. However, other

Authors stated to focus the analysis on another barcode, the nuclear ribosomal ITS region [111] **(Table 1)**.

As regards protists, the DNA barcoding worked really well in discriminate between species. However, it is only limited by the lack of the possibility to use universal primers. The primers of Folmer worked only in the study of Nassonova and collaborators (2010) [112], on naked lobose amoebae. In the other studies it was necessary to design sets of specific primers for the different groups. However, so far the DNA barcoding tool was tested only in few groups of protists: amoebae [91, 112, 113], *blastocystis* [114], diatoms [115-119], dinoflagellates [120-122], and the ciliated protists of the genera *Paramecium* [123] and *Tetrahymena* [124, 125]. In six amoebae morphospecies, Nassonova and collaborators 2010 [112] showed that the *cox1* gene resolved inter-species relationships better than any other gene. However, the intraspecific and interspecific values didn't allow establishing a clearly defined threshold. Heger and collaborators (2010) [113] and Kosakyan and collaborators (2011) [91] applied the *cox1* gene in identify nebelid testate amoebae. The *cox1* gene not only correctly separated all the studied morphospecies, but also revealed a large number of cryptic species. Regarding *blastocystis*, Scicluna and collaborators (2006) [114] developed the DNA barcoding identification system in these parasite, but they used the nuclear ribosomal ITS region as barcode. Also Moniz and Kaczmarska (2009) successfully utilized the nuclear ribosomal ITS region to discriminate diatoms species [118]. Evans and collaborators (2007) gave the first evidence of the applicability of the *cox1* gene in discriminate diatoms species [115]. Hamsher and collaborators (2011) [119] proposed an alternative DNA barcode for diatoms species identification, the large subunit of rubisco (*rbcL*-3P). Stern and collaborators (2010) [122] used the *cox1* gene as barcode for a large-scale environmental study on dinoflagellates, revealing a higher diversity in that microorganisms than expected. In dinoflagellates, Lin and collaborators (2009) [120] showed that a 385 bp of the cytochrome b (*cob*) gene had a more powerful resolution than *cox1*. In *Paramecium*, Barth and collaborators (2006) [123] showed that the *cox1* gene revealed higher intraspecific variability than the nuclear 18S gene and the ITS region. In *Tetrahymena*, Chantangsi and collaborators (2007) [126] clearly highlighted the feasibility of establishing a *Tetrahymena* species identification system reliant on the analysis of the sequence of the *cox1* gene **(Table 2)**.

Table 1. Different marker (or barcode) for different organisms.

ORGANISM	MARKER
METAZOANS	*cox1*
PLANTS	*matK* + *rbcL*
FUNGI	ITS
PROTISTS	*cox1*

Table 2. Examples of *cox1* successes. The table also highlights the method used to establish the efficacy of the *cox1* gene.

ORGANISM	METHODS	COMMENTS
AMOEBAE	ML; BI	Found cryptic diversity. Lack of a clearly defined barcoding gap.
NORTH AMERICA BIRDS	NJ; threshold	
Paramecium	NJ	
Tetrahymena	NJ; threshold	Found cryptic diversity.
TROPICAL LEPIDOPTERA	NJ; threshold	

2.3 **PROTISTS: GENERAL FEATURES**

Protists are unicellular eukaryotic microorganisms with sizes ranging from 2 to 2000 µm. In some cases they form multicellular colonies, without specialized tissues. They are lively and fascinating organisms, showing a gargantuan variety for all they feature: size, shape, symmetries, morphology, habitat, behaviour, biochemistry, physiology, metabolism, reproduction, locomotion, and lifestyle. Above all, they are incredibly rich in species. To date more than 200000 species of protists have been described [127]. Accordingly, the phylogenetic breadth of protists far exceeds that of the multicellular organisms, and this is the problem in studying them [128-132].

Protists play a significant role in ecology, health, and biotechnology [127]. Consequently, they also have a significant role in economy. They are major constituents of benthic and planktonic communities, occurring in terrestrial, marine and freshwater environments worldwide. They play a key role in the functioning of many ecosystems. They are used as bioindicators of water quality and past climates. Protists can be both free-living and symbiotic. Free-living protists can be photosynthetic or heterotrophic. Photosynthetic protists are primary producers in many trophic chains. Protists are the protagonist of essential microbial loops, affecting the health and survival of higher animals, humans included. For example, they are the basis of the ecosystem of coral reefs. Non-photosynthetic (or heterotrophic) protists are the protagonist of all nutrient cycling. Protists establish several symbiotic relationships, with a wide range of hosts, and some species are parasitic. Symbiotic species protists are 15% of all estimated number of protists species, and occur within several protistan groups. Some protists are invasive species responsible of significant loss of biodiversity. Every year hundreds of millions of dollars are thrown away due to huge losses in farm production. Also fishes and fungi are decimated every year by parasitic protists. On the opposite side, helpful symbiotic protists actively control bacteria overgrowth in nature. Some symbiotic protists ensure the health of their hosts (humans included), because they live in their digestive tracts.

Protists play an important role in health because they are pathogens of humans too. They are pathogens of diseases with the highest mortality rates in the world, for example malaria (*Plasmodium*). Other examples of serious human diseases caused by protists are toxoplasmosis (*Toxoplasma*) and leishmaniasis (*Leishmania*).

Protist involments in food, medical, and industrial products are many. Cervia and collaborators (2006, 2007) [1, 2] demonstrated that the ciliated protozoan *Euplotes crassus* produce a secondary metabolite, "euplotin C", that has a cytotoxic and pro-apoptotic action in tumor cell lines. These studies suggested the feasibility of production of new antitumor drugs from this ciliated protozoan. Fossil protists (foraminiferans) are used to determine the ages of lands to find petroleum.

Most important, protists have a great evolutionary importance. In fact, they are the first eukaryotic organisms that appeared on the Earth about two billion years ago. They are the protagonists of the biggest evolutionary step in the life story, the transition from prokaryotic cell organization to that eukaryotic, from which pluricellular organisms evolved.

The major aspects of protist biology are poorly understood and consequently contentious, for example their biodiversity and geographical distributions. The biodiversity of protists in nature is the most underestimated. The diversity of protists will be described in a comprehensive way when all geographic locations will have been sampled. The biogeography of protist is another hard issue. It is a very actual subject in scientific literature and conferences [133]. It is also a subject of debate. Some scientists argued that protists have a defined and limited geographic dispersal (endemic hypothesis) [134], while other Authors stated that protists have a global dispersion (cosmopolitan hypothesis) [135]. It is difficult to know the biogeography of protists, as not all regions have been sampled and most regions and habitats are insufficiently sampled. Furthermore, this difficulty is also due to the fact that the taxonomic resolution at the species level in protists is still unclear. Thus the rate of discovery of new species from environmental samples remains high. Indeed, most soil, freshwater, or marine samples collected contain a multitude of undescribed species [127, 136-138]. Furthermore, a methodology for studying protists on a large scale in their natural environment is still lacking [132].

2.3.1 PROTISTS AND THEIR TANGLED TAXONOMY

Protists, despite their importance, remain in the shadows of Science and Education. In accordance with Adl and collaborators (2005), this is due to the difficulty of classifying them and identifying them. The oldest classification scheme is that of Büschli (1880–1889), which divided Protozoa into Sarcodina (amoeboid organisms), Sporozoa (a parasitic group), Mastigophora (flagellated species), and Infusoria (ciliates). This classification system is based purely on morphology. It was abandoned decades ago by protistologists, but it is, unfortunately, still used by non-protistologists, thus generating a lot of confusion. Protists were until recently considered a Kingdom apart, the Protistan Kingdom. To date, the final proposed classification of eukaryotes is that of Adl and collaborators (2005) [129]. The purpose of these Authors was to provide a new classification system stable over time and easy to update. This classification reflects our current understanding of the evolution of protists and their phylogenetic relationships. To the contrary, all the previous classification schemes took account only of the morphology. Advancing with the study of molecular phylogeny, we assist in increasing discrepancies between morphology and molecular data. Before reaching the last classification, the name of many protist groups and genera have been changed so many times that still today it is difficult to determine which name applies, and homonyms are common. Many traditional groups are no longer valid and have been abandoned. Unfortunately, even now the problem of delimiting species borderlines for many protistan types remains [139].

At the beginning, the species of protists were identified and described solely on the basis of morphological criteria. The morphological species concept has been the most widely used since the discovery of protists. Now it is widely accepted that the morphology alone is not able to recognize all species. This is mainly due to the fact that many protistan species are morphologically indistinguishable, and so misdiagnoses are common. Moreover, we already lack an accessible storage of the samples with which to make comparisons. Today, it is well known that protistan species should be defined based on a combination of morphology, DNA sequences, physiology and ecology. Molecular biology techniques have given a new dimension to the study of protists. They have allowed us to discover that their diversity and abundance in nature are much larger than expected [132]. Furthermore molecular tools provided new insights into prostist classification. DNA-based studies often showed that traditional taxonomy underestimates diversity of both

macroscopic and microscopic organisms. The expectation is that the amount of cryptic diversity (i.e. genetic diversity that is not reflected in observable morphological features) in protists is very high. The most commonly used DNA sequences for phylogenetic reconstruction of protist groups are the mitochondrial 16S ribosomal RNA (16S rRNA) gene, the nuclear ITS region and the small subunit ribosomal RNA (18S rRNA) gene.

However, still today there is a lack of a fast, efficient, inexpensive and large-scale identification tool of protists.

2.4 CILIATED PROTIST: GENERAL FEATURES

Ciliate is the common name assigned to a protist taxon comprising the phylum Ciliophora. Over 8000 ciliate species have been described from marine, freshwater and terrestrial habitats, where they play a crucial ecological role. Ciliated protists are the most abundant phagotrophs in the biosphere. They are main actors of nutrient recycling in the ecosystems. They are distinguished on morphological basis and characterized by a nuclear dualism and body-covering cilia used for both locomotion and feeding. Ciliated protists also exhibit complex behaviours, for example predation, formation of cysts and spores, and chemotaxis. Some ciliated protozoan lines are "immortal": they rejuvenate after sexual reproduction. In ethological researches they have been subjected to more study than any others protistan groups.

Aerobic ciliates also have the mitochondrial genome too. In GenBank databases only six mitochondrial genomes, all belonging to the genera *Paramecium* and *Tetrahymena*, were available before 2009. In 2009 de Graaf and collaborators [40] published the completed sequencing of the mitochondrial genomes of two *Euplotes* species, *E. crassus* and *E. minuta*. These Authors discovered that the linear mitochondrion genomes of these two *Euplotes* species show the same gene content of the other sequenced mitochondrial genomes of ciliated protozoa, but the gene order is completely different. Furthermore, these *Euplotes* species showed very long *cox* genes. However, there was no gene order difference between *E. crassus* and *E. minuta*.

Ciliated protozoa have been well known in literature for years as difficult organisms to classify, mainly because of the fact that "morphological conservationism is often associated with great ecological and molecular diversity": cryptic species [140]. Over the past 10 years the classification of ciliates has been revised several times due to new information arising from phylogenetic analysis based on molecular markers (rRNA). Taken together, the new molecular data are often at odds with the morphological data and revealed unexpected complexity and diversity.

2.4.1 THE CILIATED PROTIST GENUS *Euplotes* AND ITS TAXONOMIC IMPEDIMENT

The hypotrichous (cirri on functional ventral surface) genus *Euplotes* is the most cosmopolitan, ubiquitous, and specious group of ciliated protists. The genus *Euplotes* is defined on morphological basis. *Euplotes* has broadly adapted all over the face of the earth. There are about 100 described species but the rate of discovery of news species still remains high. Due to its abundance in nature and in its easiness of maintaining *Euplotes* strains in laboratory, the literature regarding this genus is extensive. Despite this, *Euplotes* is the most tangled taxonomy genus of ciliated protist, and its phylogenetic relationships were revised many times and they are still unclear. The identification of the *Euplotes* species requires expert skills and a lot of time. This genus, in fact, comprises of a bewildering array of sizes, morphologies, life histories, environmental tolerances and behaviours. The species identification problem in the genus *Euplotes* is mainly due to the fact that many of these species are morphologically indistinguishable. Species with similar morphologies may differ greatly with respect to other criteria. The most striking example of the problem concerning the identification of species of *Euplotes* is that of the "*E. vannus-crassus-minuta* complex". Three closely related morphotypes of *E. vannus*, *E. crassus* and *E. minuta* were indistinguishable from each others since 1980's. In literature we have been witnessing a great debate since the 1980's to answer the question of whether the three are the same species or evolutionarily distinct entities. Various Authors, basing their conclusions on morphological, isoenzymatic, mating types and ecological features, were on the side of yes or no, with neither reaching a unanimous conclusion. The final demonstration that they have three distinct evolution life histories was given by the 18S rRNA gene [141].

Traditionally, also the species of *Euplotes* were described purely on the basis of external morphology **(Fig. 5)**. In 1960, Tuffrau used numbers and distribution of cirri (compound ciliary organelles), shape of the macronucleus (vegetative nucleus), the dorsal agyrome patterns, the dorsal kineties, and the sylverline system. But the variability of characters based on the dorsal agyrome and numbers of cirri soon appeared insufficient for species identification. Borror (1972) revised the classification proposed by Tuffrau. Carter, in the same year added additional morphological features to species recognition, the number and shape of adoral zone of membranelles. The subsequent *Euplotes* revisions are those of Curds (1975), and Gates and Curds (1979), which always based their

classification scheme on morphological features. Borrow and Hill (1995), were the first to propose a new classification scheme of the *Euplotes* genus based on morphological features integrated with ecological features too.

Today it is well known that morphology alone is not adequate to distinguish *Euplotes* species. Di Giuseppe and collaborators (2011) [142] in a recently published paper, demonstrated that different *Euplotes* populations from the two poles of our planet all belonged to the *E. nobilii* morphospecies based on a combination of morphometric, genetic, phylogenetic and cross-breeding experiments. Furthermore, in the last two decades detailed investigation of secondary metabolites from the genus *Euplotes* has added “a new dimension” to the problem of their species-specific allocation via a "chemotaxonomic" approach, which is able to define protistan taxonomy to the sub specific level [3]. In fact, the huge degree of biodiversity that characterize the *Euplotes* genus is also highlighted by the secondary metabolites production. For example, the secondary metabolite “euplotin C” is produced only by the *E. crassus* morphospecies, thus enabling its identification. The other two members of the *Euplotes vannus-crassus-minuta* complex, *E. minuta* and *E. vannus*, produced only a precursor or did not produce any euplotin secondary metabolite respectively. *E. vannus*, in fact, produced a different set of secondary metabolites, which can be used to identify this species.

It would be important to have a reliable and rapid tool to distinguish *Euplotes* morphospecies because of their ecological, biochemical, and biotechnological relevance. We still lack suitable keys to enable *Euplotes* identification and little progress can be made with the purely morphological approach, for the reasons discussed so far.

Therefore, given these difficulties, why not try the DNA barcoding tool to identify *Euplotes*?

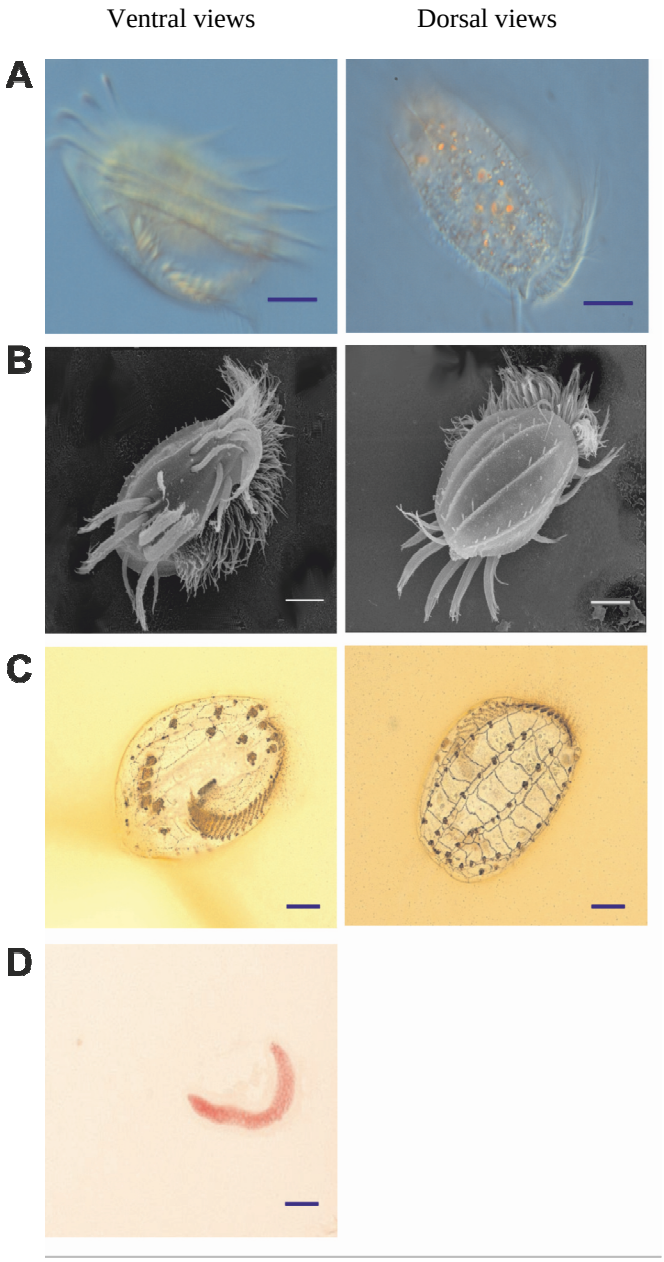


Fig. 5. (A) Differential interference contrast microscope (DIC) of *Euplotes*. **(B)** Scanning electron micrographs (SEM) of *Euplotes*. **(C)** Silver nitrate impregnation of *Euplotes*. **(D)** The staining method of Feulgen highlights the macronuclear shape of *Euplotes*. Bars 10 µm.

3 AIM OF THIS WORK

This Doctoral thesis addressed two issues.

The major purpose of this research project was to determine whether the *cox1* barcoding is an effective identification tool capable of differentiating closely related *Euplotes* species. This evaluation was conducted with two approaches. The first approach was to analyze the relationship between the interspecific and intraspecific *cox1* sequence variation. The second approach was to verify that strains belonging to the same morphospecies form a monophyletic cluster on phylogenetic trees.

The second issue was to evaluate the ability of the *cox1* gene as a phylogenetic marker. To make this evaluation, the *cox1* phylogeny was compared to the 18S phylogeny, the nuclear non-coding gene normally used to infer the phylogeny in eukaryotic organisms. The 18S gene sequences for different *Euplotes* strains and species were already available in the laboratory where I conducted my doctoral thesis.

4 MATERIALS AND METHODS

4.1 SOURCE OF SAMPLES AND CULTURE CONDITIONS

The availability of the largest collection of *Euplotes* living strains which currently exists worldwide, the result of half a century of sampling at all latitudes of the globe, allowed me to help create the basis for verifying the effectiveness of the DNA barcoding tool to identify and discriminate the *Euplotes* protist taxa.

To test the feasibility of establishing a *Euplotes* species identification system, 81 strains belonging to 15 *Euplotes* morphospecies were analyzed. **Table 3** shows the 81 *Euplotes* strains analyzed in this study and their distribution among the *Euplotes* morphospecies. For a more detailed list of the *Euplotes* strains analyzed, see **Table 4**. The **Fig. 6** shows the huge geographical dispersal of the *Euplotes* strains analyzed. All the *Euplotes* strains collected were established starting from single naturally occurring cells to obtain cellular line clones and were fed on both microalgae and bacteria, which represent a component of the natural diet of the species concerned. Marine strains were grown in salt water (32‰ salinity) on sterilized, defined and artificial seawater prepared according to Allen's formula[1] and inoculated with both *Dunaliella tertiolecta* (*Chlorophyceae*, *Dunaliellales*) (algae) and *Enterobacter aerogenes* (bacteria), grown in the same artificial seawater as ciliate organism. The freshwater *Euplotes* strains were grown in S.M.C. medium[2] and the freshwater micro-alga *Chlorogonium sp.* and the bacterium *Enterobacter aerogenes* were grown both in S.M.C. medium and were used as food organism. The S.M.C. medium was prepared from S.M.B. medium[3]. Both *Dunaliella tertiolecta* and *Chlorogonium sp.* were cultivated in aerated 5-liter Erlenmeyer flasks at 23 ± 1 °C in a 12 h light/dark cycle for at least 10 days using a daylight (Osram Daylight lamp, 36 W/10) and fluorescent (Osram Fluora lamp, 40 W/77) illuminated incubator system. Bacterial cultures were grown monoxenically either with 0.05 (w/v) protease-peptone in salt water at 4‰ for marine *Euplotes* species or distilled water for freshwater *Euplotes*, and incubated in the dark at 37 °C for 24 h. The food was added every other day in an amount that kept the cells well supplied but nevertheless avoided overfeeding (overfeeding in *Euplotes* easily

[1] 1M stock of NaCl, KCl, $CaCl_2$, $MgCl_2$, $MgSO_4$ and $NaHCO_3$

[2] S.M.C. medium (S.M.B. medium + 0.5 µM $MnCl_2$, 9.0 µM $FeCl_3$, 1.25 mM NH_4NO_3)

[3] S.M.B. medium (1.5 M NaCl, 0.05 M KCl, 0.05M $MgSO_4$, 0.4 M $CaCl_2(H_2O)_2$, 0.05 M $MgCl_2(H_2O)_6$, 0.2 M NaH_2PO_4(H2O), 0.2 M $Na_2HPO_4(H_2O)_{12}$)

slows down the multiplication rate). For a few days before the extraction of the DNA, *Euplotes* cultures were fed only with bacteria.

Normally, *Euplotes* strains were maintained in culture at 23 ± 1 °C. To the contrary, the Arctic and Antarctic *Euplotes* strains, were cultivated at 6 ± 1°C. Both bacteria and algae were kept in the fridge at 6 °C before feeding the polar *Euplotes* strains, in order to avoid an increase in temperature of these strains. The polar *Euplotes* species are: *E. euryhalinus*, *E. focardii*, *E. nobilii*, and *E. polaris*.

Table 3. Number of strains analyzed in this study for each *Euplotes* morphospecies. The number in parentheses corresponds to the number indicating the species in **Fig. 6**.

SPECIES	NUMBER OF STRAINS ANALYZED
E. crassus (1)	17
E. daidaleos (2)	3
E. euryhalinus (3)	5
E. focardii (4)	2
E. gracilis (5)	3
E. harpa (6)	7
E. magnicirratus (7)	3
E. minuta (8)	5
E. nobilii (9)	11
E. polaris (10)	4
E. quinquecarinatus (11)	5
E. raikovi (12)	5
E. rariseta (13)	4
E. vannus (14)	3
E. woodruffi (15)	4

Table 4. Detailed list of the *Euplotes* strains examined in this study with their geographical origin. The number in parentheses corresponds to the number indicating the species in **Fig. 6**.

SPECIES	STRAIN	HABITAT	ORIGIN
E. crassus (1)	Beirut3*4	Marine	Nahr Ibrahim, Lebanon
	D2		Livorno, Italy
	EC4		Porto Vecchio, France
	GCH-I1B		Galapagos Island, Ecuador
	GRG4		Galapagos Island, Ecuador
	Kuwait/4CCAP		Kuwait
	MAL1		Malindi, Kenya
	OM3E3		Norddeich, Germany
	PLR1		Canary Islands, Spain
	Oyster3*5		Cap Code, USA
	PMEX13*1S		Tulum, Mexico
	POR7		Porto Recanati, Italy
	SAFR*4		Cape Town, South Africa
	SF8		Sciacca, Italy
	Shane5		Peggy's Cove, Canada
	SL2		Bentota, Sri Lanka
	ST		San Terenzo, Italy
E. daidaleos (2)	1RP2001	Freshwater	Saint Petersburg, Russia
	PR-1		Patapsco River, USA
	SC5		Krakow, Poland
E. euryhalinus (3)	3bILb3	Freshwater, brackish water, and marine	Ilulissat, Greenland
	4-11ILb5		Ilulissat, Greenland
	ADC1		Adelie Cove, Antarctica
	LAP1		Svalbard, Norge
	WB11		Terra Nova Bay, Antarctica
E. focardii (4)	95	Marine	Terra Nova Bay, Antarctica
	MixII		Terra Nova Bay, Antarctica

SPECIES	STRAIN	HABITAT	ORIGIN
E. gracilis (5)	ATCC50191	Marine	The American Type Culture Collection
	GNI1		Sicily, Italy
	GPO4(2)		San Terenzo, Italy
E. harpa (6)	ALM10	Freshwater	Mughsayl, Oman
	ALM3		Mughsayl, Oman
	BaI5		Island of Barra, Scotland
	GS-4		Okinawa Prefecture, Japan
	Nap3		Naples, Italy
	RB22		Berezovi Island, Russia
	Serch1		Pisa, Italy
E. magnicirratus (7)	3SC10	Marine	Shelkovo City, Russia
	CO		Colombia
	Miya3		Miyajima, Japan
E. minuta (8)	CCAP1624/13	Marine	The American Type Culture Collection
	GAR4		Gerachico, Spain
	IM2		Imperia, Italy
	Kling4		Emerald Isle, USA
	MAR11		Maratea, Italy
E. nobilii (9)	1QAA	Marine	Qaanaaq, Denmark
	1QN1		Siorapaluk, Denmark
	2QAN1		Thule, Denmark
	3QAN7		Thule, Denmark
	3QN2		Siorapaluk, Denmark
	4Pyrm4		Pyramiden, Norge
	5QAA15		Qaanaaq, Denmark
	EurhB		Edmonson Point, Antarctica
	PNA-1h		Tierra del Fuego, Argentina
	QAN1		Thule, Denmark
	Sop		Icaro Faraglione Campus, Antarctica

SPECIES	STRAIN	HABITAT	ORIGIN
E. polaris (10)	EdPoB02	Marine	Terra Nova Bay, Antarctica
	Eup.Scon		Ilulissat, Greenland
	Ila7policlonale		Ilulissat, Greenland
	Split3		Terra Nova Bay, Antarctica
E. quinquecarinatus (11)	DAK5	Marine	Dakar, Senegal
	GBS-I3		Galapagos, Ecuador
	GBS3		Galapagos, Ecuador
	MR13		Sharm el Sheik, Egypt
	OMAN2		Mughsayl, Oman
E. raikovi (12)	4MNM	Marine	Manly-Wynnum, Australia
	Biod2		Biodola, Italy
	GA8		Gaeta, Italy
	LPSA5		Lampedusa, Italy
	Myy1		Manly-Wynnum, Australia
	PCE1		Porto Cesareo, Italy
	SMAaj3		Santa Maria Island, Portugal
E. rariseta (13)	BR1	Marine	Ubatuba, Brazil
	FSl1		Pisa, Italy
	MAM30		Wynnum, Australia
	Smund2		Swakopmund, Namibia
E. vannus (14)	Bali6	Marine	Bali, Indonesia
	SML		S.Maria di Leuca, Italy
	TB6		Tanabe, Japan
E. woodruffi (15)	CoMa1-5	Freshwater	Comacchio, Italy
	Fane1		Favone, France
	MS-3		Yamagata Prefecture, Japan
	SydEU6		Newcastle, Australia

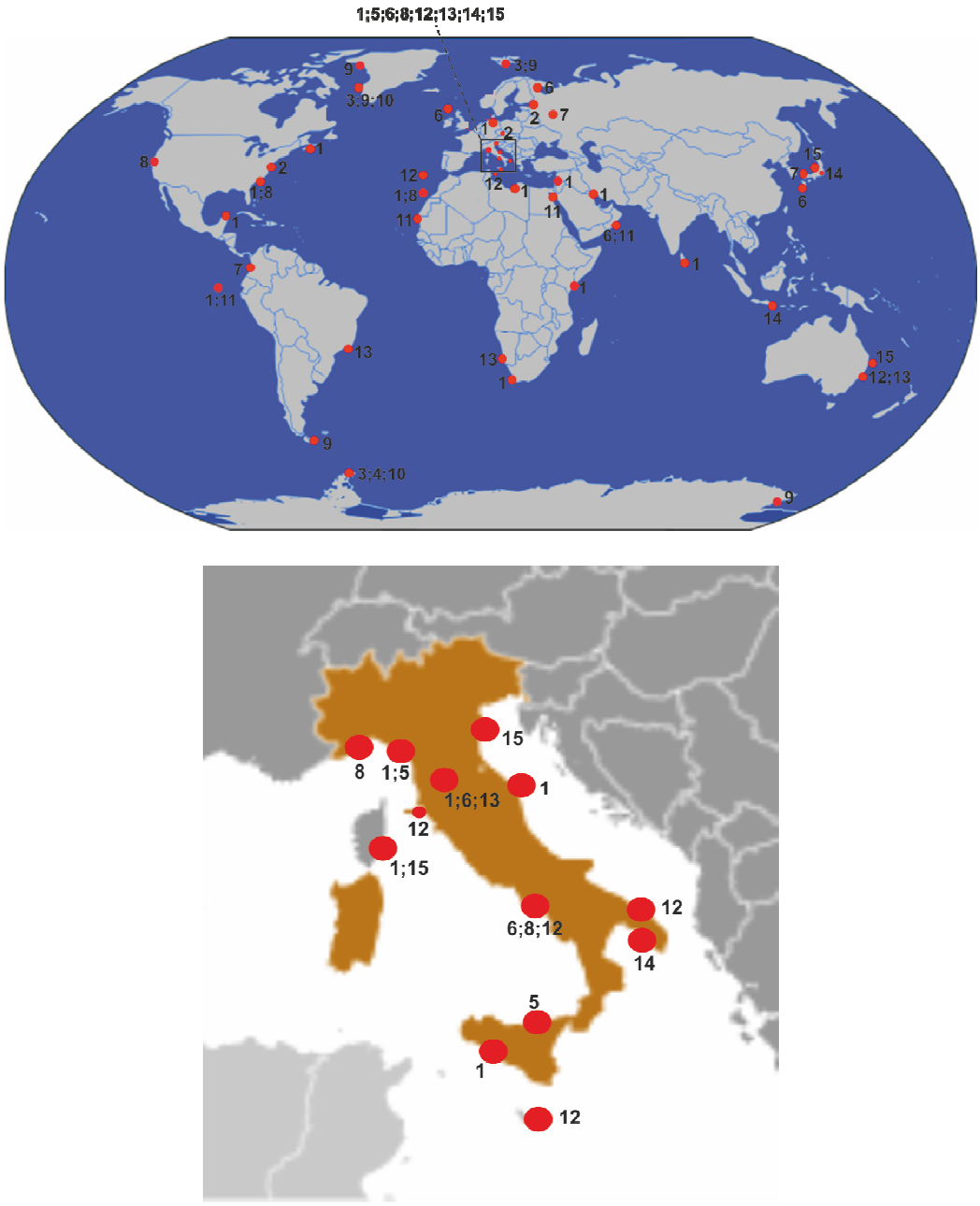


Fig. 6. The huge geographical dispersal of the sampling. Strains were collected at all latitude of the terrestrial globe. Numbers correspond to the *Euplotes* species in **Table 3** and **Table 4**. In the bottom a detail of the *Euplotes* strains collected in the Italian geographycal area and in Corse.

4.2 DNA EXTRACTION

The DNA was extracted from *Euplotes* healthy cells. I tested three different protocols for DNA extraction, selected on the basis of the initial number of cells from which to extract the DNA.

a) Guanidine protocol. This extraction protocol is used in cases where the initial number of cells is a few thousand, but less than one million. This extraction protocol is useful when it is necessary to send the sample of lysed cells, since the solution containing guanidine and lysed cells remained stable for many days even at Room Temperature (RT). I used this protocol, in fact, for the strains that I did not have available in the collection. Cells were pelleted by centrifugation and resuspended in 100 µl of distilled H_2O. 400 µl of guanidine chlorhydrate 8 M were added to cells and the sample was incubated at RT for 1 hour. Then, 400 µl of EDTA pH 8 were added and the DNA was precipitated with 0.1 volumes of LiCl 4 M and 1 volume of 100% ice-cold ethanol and preserved at -20 °C over night (O.N.), followed by centrifugation 15000 x *g* for 15 minutes at 4 °C. The pelleted nucleic acids were washed 2 times in 70% ethanol and air dried before resuspension in a suitable volume of distilled H_2O (generally 20-40 µl).

b) Phenol/Chloroform protocol. This DNA extraction protocol is performed for extraction from about one million cells. Cells were pelleted by centrifugation and resuspended in 500 µl of distilled H_2O. 150 µl of NDS (0.5 M EDTA, 1% SDS, 10 mM Tris-HCl, pH 9.5) at 55 °C and 65 µl of proteinase K (1 mg/ml H_2O) were added to cells and the sample was incubated at 55 °C for 12-15 hours. Then, 715 µl of Polyetilenglicol (PEG) 12%/NaCl 1.2 M cold were addeded and the sample was incubated on ice for 1 hour, followed by centrifugation 15000 x *g* for 15 minutes at 4 °C. Then, the pelleted nucleic acids were washed 2 times in 70% ethanol and air dried before resuspension in 500 µl of distilled H_2O. 5 µl of RNase A (10 mg/ml) were added and the sample was incubated at 37 °C for 1 hour. Then, 500 µl of phenol at 37 °C were addeded, followed by centrifugation 15000 x *g* for 2 minutes at RT. 500 µl of chloroform were added to the surnatant, followed by centrifugation 15000 x *g* for 2 minutes at RT. 400 µl of PEG 12%/NaCl 1.2 M were again added to the surnatant, and the sample was incubated on ice for 1 hour, followed by

centrifugation 15000 x *g* for 15 minutes at 4 °C. Then, the pelleted nucleic acids washed 2 times in 70% ethanol and air dried before resuspension in 400 µl of distilled H_2O. Finally, The DNA was precipitated with 0.1 volumes of LiCl 4 M and two volumes of 100% ice-cold ethanol and preserved at -20 °C for 2 hours.

c) QIAamp DNA Micro Kit (Qiagen®) protocol. This DNA extraction protocol is performed for extraction from very few cells, from 1 to about 50. According to manufacter's instruction, 80 µl of ATL Buffer were addeded to single collected cells. Then, 100 µl of AL Buffer and RNA Carrier (100 µl buffer AL + 1 µl RNA Carrier) were added. 10 µl of proteinase K (1 mg/ml) were added and the sample was incubated at 56 °C for 10 minutes. Then, 100 µl of 100% ethanol were added and the sample was incubated at RT for 3 minutes, followed by centrifugation at 6000 x *g* for 1 minute at RT. The surnatant was transferred in a QIAamp MinElute column provided by the kit followed by centrifugation at 6000 x *g* for 1 minute at RT. 500 µl of AW1 Buffer were added followed by centrifugation at 6000 x *g* for 1 minute at RT. 500 µl of AW2 Buffer were added followed by centrifugation at 6000 x *g* for 1 minute at RT. After transferring the eluate in a new tube, 50 µl of sterile H_2O were added and the sample was incubated at RT for 5 minutes, followed by centrifugation at 15000 x *g* for 1 minute at RT.

The DNA content was then spectrophotometrically determined with a DU 640 Spectrocrotometer (Beckman Instruments Inc., Fullerton, CA).

4.3 **PRIMER DEVELOPMENT, PCR AMPLIFICATION, AND SEQUENCING**

Novel and degenerates *cox1* PCR primers specific for the genus *Euplotes* were designed in this study: the forward primer 101cox1F and the reverse primer 625cox1R **(Table 5)**. These primers were designed using aligned sequences of the *cox1* gene from the only two *cox1 Euplotes* sequences available on GenBank [40] *E. crassus* (GQ903131.1) and *E. minuta* (GQ903130.1) and from available *cox1* sequences of other ciliated protistan genera *Paramecium aurelia* (NC_014262.1) and *Tetrahymena pyriformis* (NC_000862.1). I used also sequences of these non *Euplotes* species to design the primers because they are the phylogenetically closest to *Euplotes* for which sequences are available in the databases. Moreover, the nucleotide positions of the diagnostic barcoding region of the study of Chantangsi and collaborators (2007) [126] (barcoding *Tetrahymena*), corresponded to the amplified fragment from the new PCR primers 101cox1F and 625cox1R.

Approximately 1100-1600 bp of the mitochondrial *cox1* gene were amplified by PCR **(Fig. 7)**. The PCR amplification was performed in a total volume of 50 µl, including 1 mM $MgCl_2$, 0.1 µM of each primer, 250 µM of each dNTP, one unit of Taq DNA polymerase (Polymed, Florence, Italy), and 100 ng of DNA. Reactions were accomplished using a GeneAmp PCR System 2400 (Applied Biosystems, Foster City, CA) that was programmed as follows: hold at 94 °C for 1 minute, 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 50 °C for 2 minutes and extension at 72 °C for 1 minute.

The PCR products corresponding to the expected size were separated by agarose gel electrophoresis **(Fig. 8)**, purified using the GenElute™ PCR Clean-Up Kit SIGMA® and cloned into *Escherichia coli* competent cells using the TOPO TA cloning kit, following the instructions of the manufacturer (Invitrogen®, Carlsbad, CA). *Escherichia coli* competent cells were transformed by the cloning vector (plasmide) that has linked inside the amplified PCR fragment, thus giving rise to recombinant colonies.

The plasmidic DNA was extracted from recombinant colonies by lising of the cells with the non-ionic and non-denaturing detergent Nonidet P 40 0.1% to carry on a PCR of control using the standard cloning vector primers M13 forward and M13 reverse **(Fig. 9** and **Table 5)**. This PCR of

control was done to verify the correct size of the DNA inserts contained within the cloning vectors. The PCR of control was performed in a total volume of 25 µl, including 1 mM $MgCl_2$, 0.1 µM of each primer, 250 µM of each dNTP, one unit of Taq DNA polymerase (Polymed), and 3 µl of plasmidic DNA. Reactions were accomplished using a GeneAmp PCR System 2400 (Applied Biosystems) that was programmed as follows: hold at 94 °C for 1 minute, 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 50 °C for 2 minutes and extension at 72 °C for 1 minute.

The PCR products were separated by agarose gel electrophoresis to verify the presence of the insert of corresponding dimensions to the amplified PCR fragment **(Fig. 10)**. The PCR products corresponding to the expected size were precipitated with 0.1 volumes of CH_3COONa 3 M pH 4.8 and two volumes of 100% ice-cold ethanol and preserved at -20 °C for 2 hours, followed by centrifugation 15000 x *g* for 15 minutes at 4 °C. Then, the pelleted nucleic acids were washed two times in 70% ethanol and air dried before resuspension in 40 µl of distilled H_2O. The products of the PCR of control were purified using the GenElute™ PCR Clean-Up Kit SIGMA®.

The purified products of the PCR of control were used to carry out the PCR of sequencing using the cloning vector primers **(Fig. 11)**. The PCR of sequencing was performed in a total volume of 10 µl, including the reaction Mix comprising the four nucleotides labeled with different fluorochromes, and the products of the PCR of control. Reactions were accomplished using a GeneAmp PCR System 2400 (Applied Biosystems) that was programmed as follows: hold at 94 °C for 1 minute, 30 cycles of denaturation at 94 °C for 30 seconds, annealing and extension combined together in a single step at 60 °C for 4 minutes.

The sequences thus obtained were used to draw a set of specific internal primers for each species to direct sequencing. This was necessary because of the very high nucleotidic variability of the tested species. In some cases, it was necessary to draw a set of internal primers specific to single *Euplotes* strains.

The 18S gene sequences for all *Euplotes* strains and species tested for the *cox1* gene were already available in the laboratory where I conducted my doctoral thesis.

Table 5. Sequences of the degenerated[4] *cox1* and of the cloning vector primers.

NAME	SEQUENCE
101cox1F	5'-ACIGCICAYGGIBTIATHATGG-3'
625cox1R	5'-CATIGGRTTIGCIGCDATCATC-3'
M13 forward	5'- GTAAAACGACGGCCAG-3'
M13 reverse	5'-AGGAAACCAGCCCTAATTTGGACC-3'

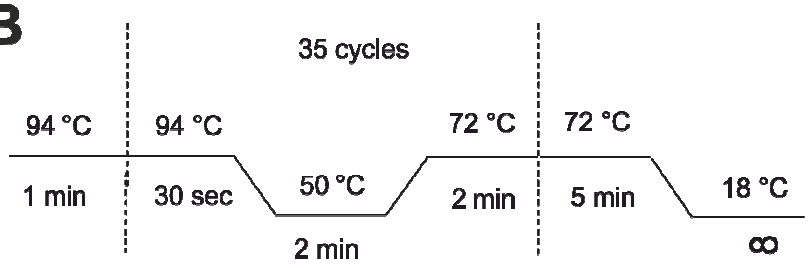
[4] Y = C or T; B = C, G or T; H = A, C or T; R = A or G; D = A, G or T.

PCR

A

H_2O	34 µl
PCR Buffer 10X	5 µl
Forward Primer (0.1 µg/µl)	1 µl
Reverse Primer (0.1 µg/µl)	1 µl
dNTP (2.5 mM)	5 µl
$MgCl_2$ (50 mM)	2 µl
DNA (100 ng/µl)	1 µl
Taq (1U/µl)	1 µl
Total volume	50 µl

B



C

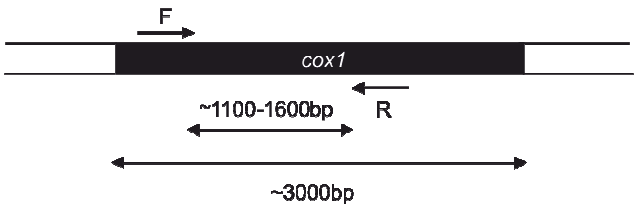


Fig. 7. PCR amplification of the *cox1* gene. **(A)** The reagents. **(B)** The amplification protocol. **(C)** The position of the primers with respect to the total length of the gene.

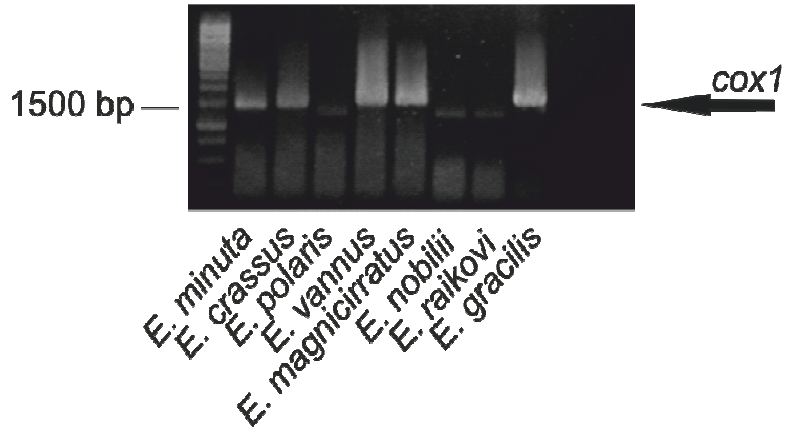


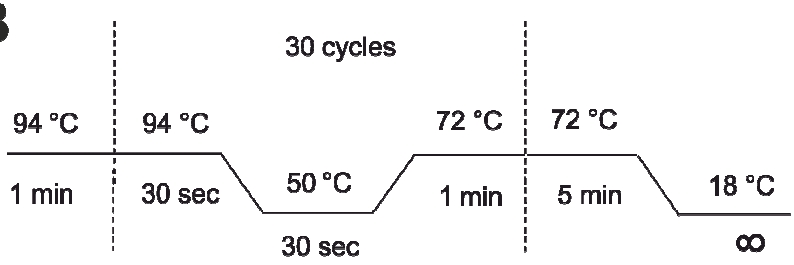
Fig. 8. Amplified *cox1* fragments by PCR.

A

PCR of CONTROL

H_2O	12 µl
PCR Buffer 10X	2.5 µl
M13Forw Primer (0.1 µg/µl)	1.5 µl
M13Rev Primer (0.1 µg/µl)	1.5 µl
dNTP (2.5 mM)	2.5 µl
$MgCl_2$ (50 mM)	1 µl
DNA	3 µl
Taq (1U/µl)	1 µl
Total volume	25 µl

B



C

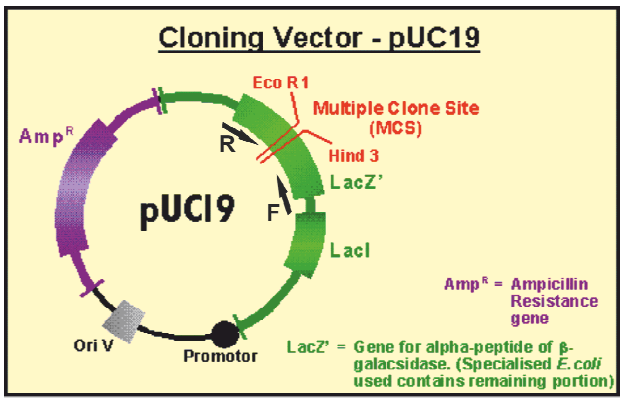


Fig. 9. The PCR of control. **(A)** The reagents. **(B)** The amplification protocol. **(C)** Specific positions of the primers M13 Forward (F) and M13 Reverse (R) next to the insert zone of the cloning vector.

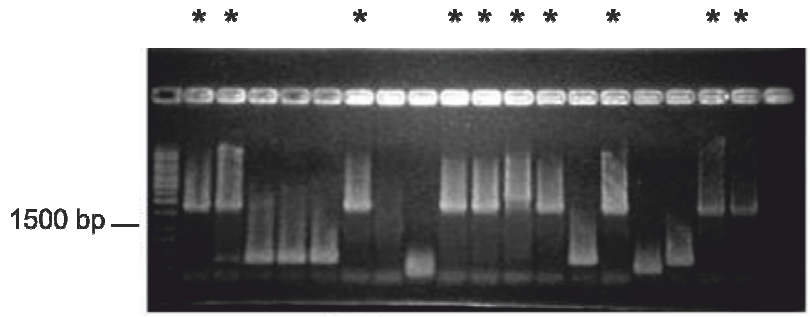


Fig. 10. Gel electrophoresis of the products of the PCR of control. The symbol "*" indicates the recombinant clones containing the insert of corresponding dimensions to the amplified PCR fragment.

A PCR of SEQUENCING

Mix	4 µl
Primers forward (0.04 µg/µl)	2 µl
Primers reverse (0.04 µg/µl)	2 µl
DNA	4 µl
Total volume	10 µl

B

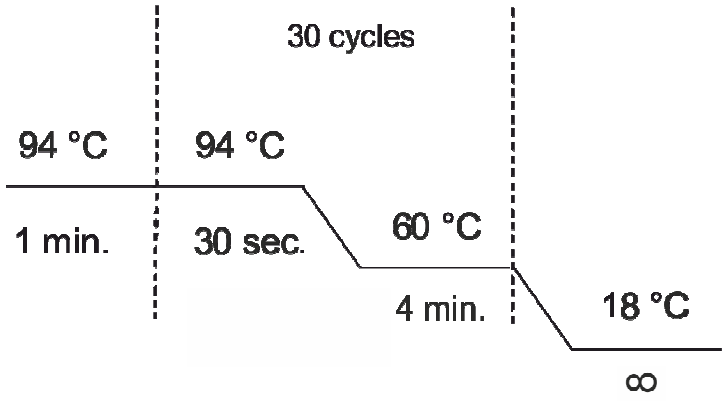


Fig. 11. The PCR of sequencing. **(A)** The reagents. **(B)** The PCR protocol.

4.4 MULTIPLE SEQUENCE ALIGNMENT (ALN)

The problem of constructing an accurate alignment is often overlooked. The construction of an accurate ALN is the crucial point of phylogenetic analysis, because all subsequent analysis is based on this ALN. The MAFFT program is one of the most accurate programs [143, 144]. It is very important to point out that MAFFT doesn't insert gaps inside the codons, an error that is well known to lead to wrong phylogenetic reconstruction.

The multiple alignment of the *cox1* gene sequences was produced using the pipeline implemented in the TranslatorX server [145]. This web-based tool allows the aligning of orthologous nucleotide sequences using as a backbone the alignment obtained for the corresponding translated polypeptides. The MAFFT program was used to align the amino acid data set in the TranslatorX server pipeline. The "The Mold, protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma / Spiroplasma Code (table = 4)", available in databases (GenBank/EMBL) were used as genetic code.

The sequences thus aligned where subjected to the analysis by using the MEGA5 program [146] and the K2P [50] model of sequence evolution.

The chi-square test, that compares the nucleotide composition of each sequence to the frequency distribution assumed in the maximum likelihood model, was implemented in TREE-PUZZLE 5.2 [147].

The alignment of the 18S gene sequences from the same strains analyzed for *cox1* gene was produced by the MAFFT program.

4.5 TESTING FOR MUTATIONAL SATURATION

The accuracy of phylogenetic reconstruction depends not only on the accuracy of ALN reconstruction, but also on sequence divergence. Too much diverged sequence can lead to wrong phylogenetic reconstruction due to the problem of substitution saturation. In general, in coding genes the third codon positions evolve much faster than the first and the second codon positions. As a consequence, it is usual for third codon positions to be saturated. Generally, to avoid this problem, third codon positions are excluded from the analysis, and/or amino acidic sequences are analyzed.

An a priori estimation of the phylogenetic signal present in the multiple alignments was performed by maximum likelihood mapping [148]. The phylogenetic signal was evaluated using the TREE-PUZZLE 5.2 program [147].

The level of mutational saturation was estimated by plotting ML-estimated distances, *i.e.* the General Time Reversible plus Gamma, (GTR+G), against uncorrected p-distances (based on observed substitutions) for each multiple alignment. After fitting a regression line, its slope (R^2) was used as a measure of mutational saturation. If $R^2 = 1$ no saturation is inferred, while for $R^2 < 1$, the phylogenetic signal is largely saturated. [149].

4.6 PHYLOGENETIC ANALYSIS

Phylogenetic trees were inferred using Bayesian Inference (BI), Maximum Parsimony (MP), Maximum Likelihood (ML) and Neighbour-Joining (NJ) methods. The ciliated protists *Paramecium caudatum* (NC_014262.1) and *Tetrahymena pyriformis* (NC_000862.1) were selected as outgroups species. This decision was made on the basis that they are phylogenetically closer to *Euplotes* taxa whose sequences are available in the databases.

The BI was computed with MrBayes program [150]. The MP and NJ trees were computed with the MEGA5 program. The K2P model of sequence evolution was applied to the NJ analysis. The ML analyses was performed with the RaxML 7.2.6 program [151] implemented in the graphical user interface raxmlGUI 0.93 [152]. The GTR+G and GTR+CAT substitution models were applied to the nucleotidic alignments [153]. The reason to use the GTR+G model instead of the more popular GTR+I+G model is due to the fact that it is impossible to estimate both parameters reliably [153]. The MTREV+G model was applied to the mitochondrial encoded protein ALNs. The choice of MTREV matrix among available alternatives was done by comparing the likelihood values obtained from trees built using the different models available in RaxML 7.2.6 program.

Nonparametric bootstrap (BT) tests [154] were performed to assess the robustness of tree topologies (1000 replicates in all cases). They were taken into account only the bootstrap values higher than 50%.

I wanted to test the potential of the *cox1* gene to identify *Euplotes* species. The test was to verify if *cox1* nucleotidic sequences from the *E. crassus* (GQ903131.1) and *E. minuta* (GQ903130.1), whose mitochondrion genome have been recently made available on GenBank databases [40], were correctly clustered into appropriate species in the phylogenetic tree. For this task, I inferred the phylogeny from the *cox1* sequences available on databases together with the sequences of *E. crassus* and *E. minuta* strains that I tested in this work. Nucleotidic sequences were first aligned by ClustalW [155]. Phylogeny was inferred by the NJ method with the K2P model. I chose this method because it has been used in other DNA barcoding studies of ciliated protists [126]. The *cox1* sequences of *Euplotes* strains belonging to the species *E. quinquecarinatus* and

E. focardii were selected as outgroups, since in the 18S gene inference it has appeared that these species are sister taxa of *E. crassus* and *E. minuta* species. Non parametric bootstrap (BT) tests [154] were performed to assess the robustness of tree topologies (1000 replicates in all cases).

5 RESULTS

5.1 SEQUENCES OBTAINED

I obtained amplified fragments of different lengths depending on the tested species **(Table 6)**. The length of the amplified fragments ranged from 1167 bp in *E. raikovi* to 1602 bp in *E. magnicirratus*. Even within some *Euplotes* species I obtained amplified fragments of different lengths. For example, regarding the *E. crassus* species, I amplified by PCR fragments of 1518 bp in length, with the exceptions of the PLR1 (1482 bp) and the SL2 (1512 bp) strains. This also happened in the species *E. euryhalinus*, *E. magnicirratus*, *E. polaris*, *E. quinquecarinatus*, *E. rariseta*, *E. vannus*, and *E. woodruffi*. Regarding the *E. Euryhalinus* species, the PCR primers amplified fragments of 1491 bp in length in the 3bILb3, 4-11Ilb5, and ADC1 strains, while they amplified fragments of 1503 bp in length in the LAP1 and WB11 strains. The amplified fragment of the *E. magnicirratus* strain Miya3 was 1542 bp in length, while the amplified fragments of the conspecific strains 3SC10 and CO, were 1602 bp in length. Arctic *E. polaris* strains EdPoB02 and Split3 produced amplified fragments of 1392 bp in length, while the amplified fragments of the Antarctic *E. polaris* strains Eup.Scon and Ila7policlonale were 1401 bp in length. The *E. quinquecarinatus* strains DAK5, MR13, and OMAN2 produced PCR-fragments of 1575 bp in length, while the *E. quinquecarinatus* strains GBS3 and GBS-I3 were 1590 bp in length. Only two of the four *E. rariseta* strains showed the same length of amplified fragments. The *E. rariseta* strains Smund2 and FSl1 showed the same length (1386 bp). To the contrary, the amplified fragments of the strains BR1 and MAM30 were 1416 bp and 1422 bp in length, respectively. All the three *E. vannus* strains analyzed showed different lengths of amplified fragments. The MS-3 *E. woodruffi* strain was different in length from the other three *E. woodruffi* strains analyzed. In fact, while it was 1494 bp in length, the other three strains CoMa1-5, Fane1 and SydEu6 were 1509 bp in length.

Distance matrices of pairwise dissimilarities were performed from both the *cox1* gene and the 18S gene datasets. Very little identical nucleotidic sequences were found in the *cox1* gene dataset, while the number of identical nucleotidic sequences was much higher in the 18S gene dataset. In practice, the *cox1* gene was able to distinguish 78 haplotypes, while the 18S gene was able to distinguish 41 ribotypes.

The chi-square test showed that the *cox1* dataset is extremely heterogeneous. In **Table 7** only the strains showing p-values greater than 5% were shown. For example, the p-values for the *E. nobilii* strains were really high and ranged from 41.66% in the 4Pyrm4 strain to 76.06% in both the 1QAA and QAN1 strains. These were not the only examples of strains for which the nucleotidic sequences were highly heterogeneous. Also the *E. raikovi* strain SMAaj3 showed a really high p-value of 20.53%. This happened also in the *E. magnicirratus* strains 3SC10 (p-value = 84.30%) and CO (p-value = 44.97%). Also many *E. crassus* strains were characterized by really high p-values. They were Beirut3*4 (p-value = 29.31%), PMEX13*1S (p-value = 22.97%), POR7 (p-value = 16.30%), SAFR*4 (p-value = 12.23%), ST (p-value = 18.49%), SF8 (p-value = 23.80%), GRG4 (p-value = 9.79%), and Kuwait/4CCAP (p-value = 7.58%). Moreover, both the *E. focardii* strains analyzed showed high heterogeneity at the *cox1* nucleotidic sequences level. In fact, the strains MixII and 95 showed p-values of 5.75% and 6.35%, respectively. The GNI1 *E. gracilis* strain showed a p-value of 9.98%. Sensationally high levels of heterogeneity occurred in the *cox1* sequences corresponding to the *E. quinquecarinatus* strains GBS3 and GBS-I3, whose p-values were 71.70% and 67.98%, respectively.

The *cox1* gene is a protein-coding gene. All the amplified sequences were in frame +1 and fully functional gene sequences.

Table 6. Length of the amplified *cox1* fragments in the different *Euplotes* species analyzed.

SPECIES	LENGTH OF THE AMPLIFIED *cox1* FRAGMENTS (bp)
E. crassus	1482; 1512; 1518
E. daidaleos	1551
E. euryhalinus	1491; 1503
E. focardii	1506
E. gracilis	1566
E. harpa	1482
E. magnicirratus	1542; 1602
E. minuta	1467
E. nobilii	1179
E. polaris	1392; 1401
E. quinquecarinatus	1575; 1590
E. raikovi	1167
E. rariseta	1386; 1416; 1422
E. vannus	1542; 1545; 1551
E. woodruffi	1494; 1509

Table 7. Chi-square test (%) for some *Euplotes* strains analyzed with a p-value higher than 5%.

STRAIN	SPECIES	p-value	STRAIN	SPECIES	p-value
Beirut3*4	*E. crassus*	29.31%	1QN1	*E. nobilii*	48.75%
GRG4	*E. crassus*	9.79%	2QAN1	*E. nobilii*	39.90%
Kuwait/4CCAP	*E. crassus*	7.58%	3QAN7	*E. nobilii*	47.32%
PMEX13*1S	*E. crassus*	22.97%	3QN2	*E. nobilii*	43.89%
POR7	*E. crassus*	16.30%	4Pyrm4	*E. nobilii*	43.89%
SAFR*4	*E. crassus*	12.23%	5QAA15	*E. nobilii*	41.66%
SF8	*E. crassus*	23.80%	EurhB	*E. nobilii*	66.86%
ST	*E. crassus*	18.49%	PNA-1h	*E. nobilii*	66.86%
MixII	*E. focardii*	5.75%	Sop	*E. nobilii*	66.86%
95	*E. focardii*	6.35%	QAN1	*E. nobilii*	76.06%
GNI1	*E. gracilis*	9.98%	GBS3	*E. quinquecarinatus*	71.70%
CO	*E. magnicirratus*	44.97%	GBS-I3	*E. quinquecarinatus*	67.98%
3SC10	*E. magnicirratus*	84.30%	SMAaj3	*E. raikovi*	20.53%
1QAA	*E. nobilii*	76.06%			

5.2 NUCLEOTIDE COMPOSITION

Relative nucleotide frequencies for the *cox1* gene are shown in **Table 8**. They ranged from 32.0% in *E. minuta* to 41.0% in *E. woodruffi* for T; from 10.4% in *E. daidaleos* to 23.2% in *E. vannus* for C; from 22.7% in *E. vannus* to 36.0% in *E. daidaleos* for A; and from 9.6% in *E. daidaleos* to 21.5% in *E. vannus* for G. The C+G content was highly variable and ranged from 20.1% in *E. daidaleos* to 44.8% in *E. vannus*. Overall, the mean of the G+C content was 33.8%.

Table 8. Nucleotidic composition of the amplified *cox1* sequences in the *Euplotes* species analyzed.

SPECIES	NUCLEOTIDE FREQUENCIES (%)					
	T	C	A	G	G+C	A+T
E. crassus	34.7	20.8	27.9	16.5	37.4	62.6
E. daidaleos	43.9	10.4	36.0	9.6	20.1	79.9
E. euryhalinus	33.0	23.0	26.3	17.7	40.7	59.3
E. focardii	32.6	19.2	31.2	17.1	36.2	63.8
E. gracilis	32.5	20.7	31.0	15.7	36.5	63.5
E. harpa	40.2	11.9	33.9	14.0	25.9	74.1
E. magnicirratus	35.9	17.6	32.7	13.8	31.4	68.6
E. minuta	32.0	22.5	29.0	16.5	39.0	61.0
E. nobilii	34.7	20.2	29.3	15.8	36.0	64.0
E. polaris	32.8	18.6	35.1	13.5	32.1	67.9
E. quinquecarinatus	33.5	20.6	29.3	16.7	37.2	62.8
E. raikovi	38.5	15.9	31.7	13.9	29.8	70.2
E. rariseta	34.9	20.9	28.8	15.5	36.3	63.7
E. vannus	32.5	23.2	22.7	21.5	44.8	55.2
E. woodruffi	41.0	13.8	32.5	12.7	26.5	73.5
Mean	35.5	18.6	30.5	15.4	33.8	64.3

5.3 MULTIPLE ALIGNMENT (ALN)

The *cox1* gene sequence alignment is shown in **Fig. 12**. The alignment highlighted a huge nucleotidic divergence between the aligned sequences, especially in the central portion of the alignment.

Even the *cox1* amino acidic sequence alignment showed a high degree of variability mostly in the central portion of the alignment **(Fig. 13)**.

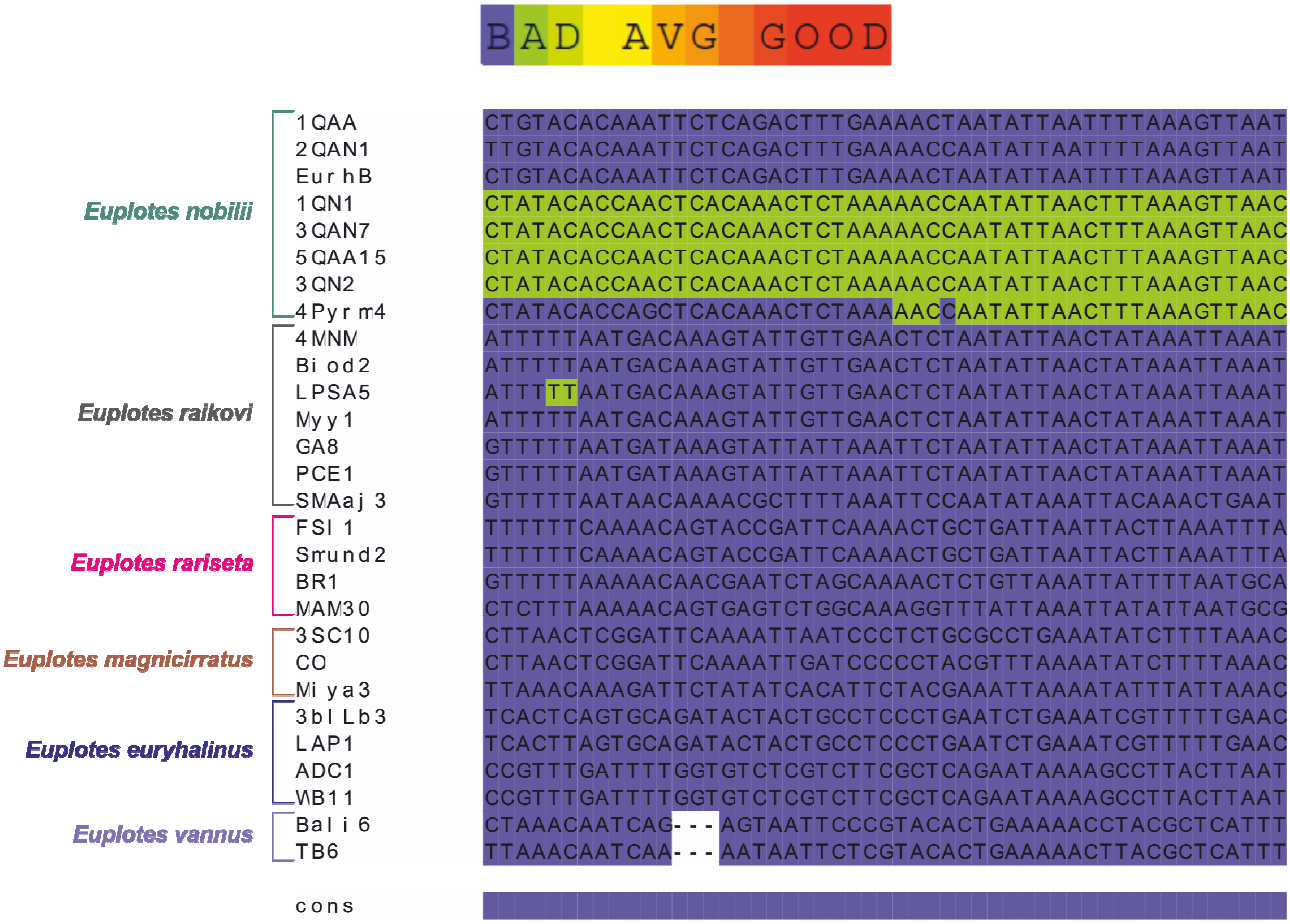


Fig. 12. A portion of the *cox1* gene sequence alignment. The color code shown in the top gave an indication of the similarity of the aligned sequences.

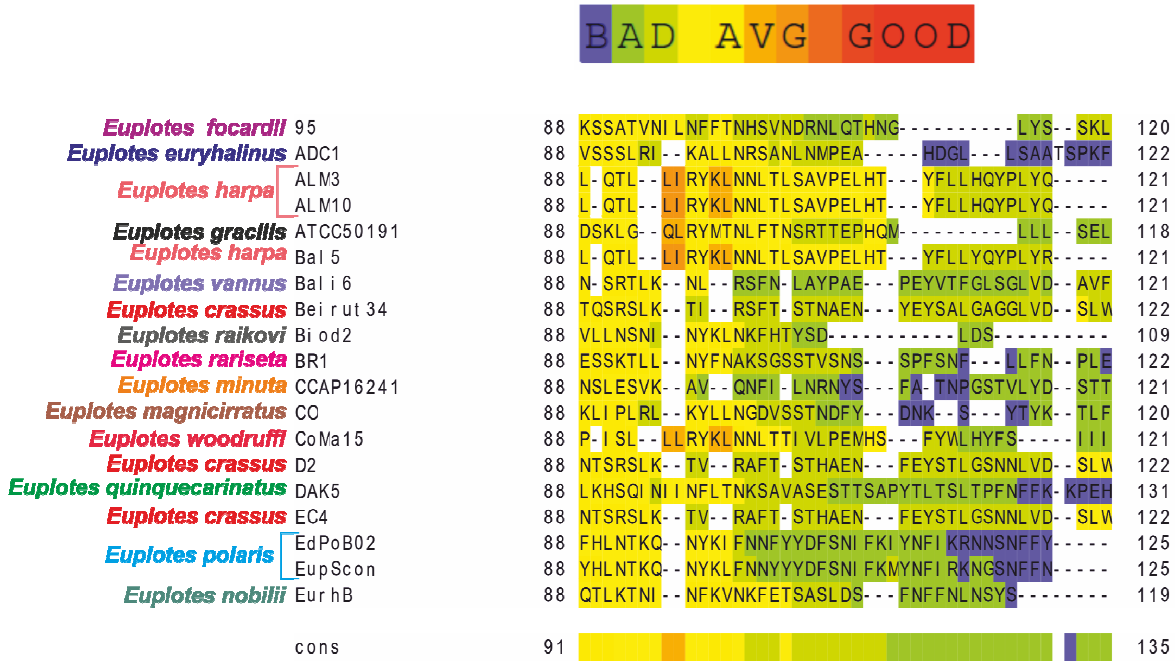


Fig. 13. A portion of the amino acidic *cox1* sequence alignment. The color code shown in the top gave an indication of the similarity of the aligned sequences.

5.4 SEQUENCE ANALYSIS OF THE *cox1* GENE

The percentages of the *cox1* nucleotidic sequence divergence between and within the *Euplotes* species were calculated using the K2P distance model of evolution **(Table 9)**. The percentage of *cox1* nucleotidic sequence divergence between species was 62.7%. The *cox1* gene discriminated very clearly between the *Euplotes vannus-crassus-minuta* complex, in fact the divergence between these three related morphospecies was 47.8%. Moreover, the variation in percentage of intraspecific *cox1* sequence divergence values was extensive and ranged from 0.5% in *E. daidaleos* to 43.7% in *E. rariseta*. Furthermore, the following seven morphospecies showed sensationally high percentages of intraspecific sequence divergence values: *E. crassus*, *E. euryhalinus*, *E. magnicirratus*, *E. polaris*, *E. quinquecarinatus*, *E. rariseta*, and *E. vannus*. However, these values were reduced by treating the groups identified within the species by the phylogenetic analysis as independent evolutionary units **(Paragraph 5.6)**. For example, *E. crassus* showed 33.6% of intraspecific sequence divergence value, and this value decreased to 25.2% when the PLR1 and SL2 strains were excluded from the analysis. In fact, the PLR1 and SL2 *E. crassus* strains formed a group apart in all the inferred phylogenetic trees **(Paragraph 5.6)**. I divided all the *E. crassus* strains into four groups, based on the phylogenetic analysis. Group 1: ST, POR7, SAFR*4, SF8, PMEX13*1S, and Beirut3*4. Group 2: MAL1, Kuwait/4CCAP, and GRG4. Group 3: Shane5, GCHI-1B, OM3E3, D2, Oyster3*5, and EC4. Group 4: PLR1 and SL2. Then, I calculated again the percentages of intraspecific sequence divergence within groups 1, 2, 3, and 4, and they were respectively: 2.8%, 15.7%, 12.4%, and 49.8%. Similarly, I considered the *E. euryhalinus* species, whose *cox1* intraspecific sequence divergence was 40.6%, as divided in two groups. The group 1, comprising of the Arctic strains 3bILb3, 4-11Ilb5, and the Antarctic strain LAP1, showed 1.2% of *cox1* intraspecific sequence divergence. Any difference was found in the sequence divergence values within the *E. euryhalinus* Antarctic strains ADC1 and WB11. Similarly, I treated the *E. magnicirratus* strains CO and 3SC10 as independent evolutionary units from the remaining conspecific strains. The *cox1* sequence divergence value of the strains CO and 3SC10 was 14.9%, versus the 40.0% when considering the overall *E. magnicirratus* strains. No differences in the sequence divergence values were found within the Arctic (Eup.Scon and Ila7policlonale) and Antarctic (EdPoB02 and Split3) *E. polaris* strains. This was true also for the *E. quinquecarinatus* strains GBS3 and GBS-I3. To the contrary, the *E. quinquecarinatus* strains DAK5, MR13, and OMAN2 showed 8.9% of *cox1* sequence divergence, and this value was three

times smaller than the overall *E. quinquecarinatus* intraspecific sequence divergence. About *E. rariseta*, the phylogenetic inference distinguished two groups. The first group contained the strains FSl1 and Smund2. The *cox1* sequence divergence value within this group was 1.8%. To the contrary, the *cox1* intraspecific sequence divergence value of the *E. rariseta* species was 43.7%. Furthermore, the BR1 and MAM30 *E. rariseta* strains showed 38.9% of *cox1* sequence divergence value. I considered the Bali6 and TB6 *E. vannus* strains as independent evolutionary units, and their *cox1* sequence divergence value was 16.9%. The intraspecific *cox1* sequence divergence value of the species *E. vannus* was much higher (35.3%).

The DNA barcoding gap, defined as the division between the percentage of interspecific and intraspecific *cox1* sequence variation, did not assume a fixed value, and it ranged from 1.4 in *E. rariseta* to 125.4 in *E. daidaleos*. However, the variability of the DNA barcoding gap values decreased if considering the new values arising from the examination of the groups within the species highlighted by the phylogenetic inference as independent evolutionary units. In fact, by doing so, the DNA barcoding gap ranged from less than 2 for the *E. crassus* group formed by the PLR1 and SL2 strains and the *E. rariseta* group formed by the BR1 and MAM30 strains, to about 50 in the group formed by the *E. euryhalinus* strains 3bILb3, 4-11ILb5 and LAP1 **(Table 10)**.

On the other hand, p-distance values between *cox1* amino acidic sequences appeared really high in particular species **(Table 11)**. They were: *E. crassus*, *E. euryhalinus*, *E. harpa*, *E. magnicirratus*, *E. minuta*, *E. polaris*, *E. quinquecarinatus*, *E. raikovi*, *E. rariseta*, and *E. vannus*. However, these values were reduced and in some cases annulled by treating the groups within the species identified by the phylogenetic analysis as independent evolutionary units **(Paragraph 5.6)**.

Table 9. Comparison of the overall and within *Euplotes cox1* divergence values with those present in literature for other organisms.

SPECIES	INTERSPECIFIC *cox1* SEQUENCE VARIATION VALUE ± SEM%	INTRASPECIFIC *cox1* SEQUENCE VARIATION VALUE ± SEM%	DNA BARCODING GAP
ANIMALS	~ 10	~ 1	~ 10
Tetrahymena	11.3	0.65	17.4
E. crassus	62.7 ± 1.8	33.6 ± 1.2	1.9
E. daidaleos		0.5 ± 0.1	125.4
E. euryhalinus		40.6 ± 2.1	1.5
E. focardii		2.0 ± 0.1	31.3
E. gracilis		15.7 ± 1.0	4.0
E. harpa		4.6 ± 0.4	13.6
E. magnicirratus		40.0 ± 1.8	1.6
E. minuta		8.4 ± 0.6	7.5
E. nobilii		9.8 ± 0.7	6.4
E. polaris		23.9 ± 1.5	2.6
E. quinquecarinatus		28.5 ± 1.3	2.2
E. raikovi		11.4 ± 0.7	5.5
E. rariseta		43.7 ± 1.7	1.4
E. vannus		35.3 ± 1.7	1.8
E. woodruffi		4.9 ± 0.5	12.8

Table 10. *cox1* sequence divergence values within *Euplotes* groups highlighted by the phylogenetic inference.

SPECIES (STRAINS)	NUCLEOTIDIC *cox1* SEQUENCES DIVERGENCE VALUE ± SEM%	DNA BARCODING GAP
E. crassus group 1 (ST, POR7, SAFR*4, SF8, PMEX13*1S, and Beirut3*4)	2.8 ± 0.8	22.3
E. crassus group 2 (MAL1, Kuwait/4CCAP, and GRG4)	15.7 ± 1.2	4.0
E. crassus group 3 (Shane5, GCHI-1B, OM3E3, D2, Oyster3*5, and EC4)	12.4 ± 0.1	5.0
E. crassus group 4 (PLR1 and SL2)	49.8 ± 2.5	1.2
E. euryhalinus group 1 (3bILb3, 4-11Ilb5, and LAP1)	1.2 ± 0.2	52.2
E. euryhalinus group 2 (ADC1 and WB11)	0.0 ± 0.0	not calculated
E. magnicirratus group (CO and 3SC10)	14.9 ± 1.1	4.2
E. polaris group 1 (Eup.Scon and Ila7policlonale)	0.0 ± 0.0	not calculated
E. polaris group 2 (EdPoB03 and Split3)	0.0 ± 0.0	not calculated

SPECIES (STRAINS)	**NUCLEOTIDIC *cox1* SEQUENCES DIVERGENCE VALUE ± SEM%**	**DNA BARCODING GAP**
E. quinquecarinatus group 1 (GBS3 and GBS-I3)	0.0 ± 0.0	not calculated
E. quinquecarinatus group 2 (DAK5, MR13 and OMAN2)	8.9 ± 0.7	7.0
E. rariseta group 1 (FSl1 and Smund2)	1.8 ± 0.4	34.8
E. rariseta group 2 (BR1 and MAM30)	38.9 ± 1.9	1.6
E. vannus group (TB6 and Bali6)	16.9 ± 1.2	3.7

Table 11. Overall and within *Euplotes* species amino acidic *cox1* sequences p-distance values.

SPECIES	AMINO ACIDIC p-distance VALUE ± SEM%
Overall	43.9 ± 1.6
E. crassus	23.3 ± 1.2
E. daidaleos	4.0 ± 0.2
E. euryhalinus	25.0 ± 1.4
E. focardii	4.0 ± 0.3
E. gracilis	8.6 ± 1.0
E. harpa	27.7 ± 1.2
E. magnicirratus	29.7 ± 1.5
E. minuta	20.4 ± 1.0
E. nobilii	2.2 ± 0.5
E. polaris	19.9 ± 1.8
E. quinquecarinatus	31.6 ± 1.4
E. raikovi	22.5 ± 1.1
E. rariseta	30.0 ± 1.7
E. vannus	23.4 ± 1.5
E. woodruffi	2.8 ± 0.6

5.5 TESTING FOR MUTATIONAL SATURATION

5.5.1 LIKELIHOOD-MAPPING ANALYSIS

The Maximum-likelihood approach can be used to study the amount of the evolutionary information contained in a data set. The analysis is based on the Maximum-likelihood values for the three possible four taxa trees. This method partitions the area of equilateral triangle into seven regions **(Fig. 14)**. The three trapezoids at the corners represent the fully resolved quartets. The three rectangles on the sides represent the partially resolved quartets. The centre of the triangle represents the fully unresolved quartets. Consequently, only the percentages at the corners are relevant.

As shown in **Fig. 14**, the dataset comprising all the codon positions of the *cox1* gene **(A)** showed 87.9% of fully resolved quartets. To the contrary, the fully unresolved quartets were 5.5%. Regarding the dataset comprising the first and the second codon positions of the *cox1* gene **(B)**, the fully resolved quartets were 87.2%, while the fully unresolved quartets were 5.5%. Regarding the dataset comprising only the third codon positions of the *cox1* gene **(C)**, the fully resolved quartets were only 65.6%, while the fully unresolved quartets were even 25.8%. Regarding the 18S gene dataset **(D)**, the fully resolved quartets were 97.5%, while the fully unresolved quartets were 1.5%.

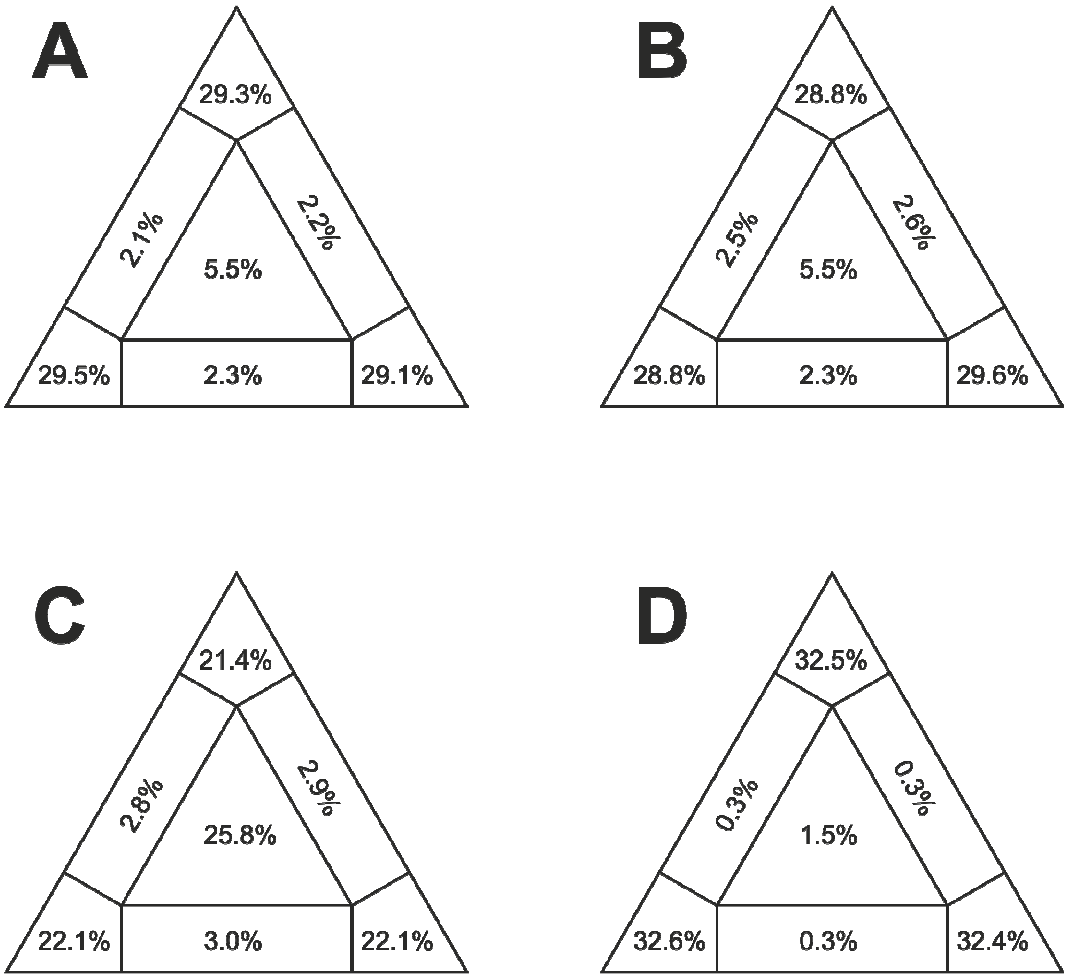


Fig. 14. (A) The equilateral triangle corresponded to the Likelihood-mapping analysis of all *cox1* codon positions. **(B)** The equilateral triangle represented the Likelihood-mapping analysis of the *cox1* gene first and the second codon positions, without the third codon positions. **(C)** The equilateral triangle showed the presence of high level of lost of phylogenetic signal in the *cox1* third codon position. **(D)** The equilateral triangle inferred from the 18S gene dataset.

5.5.2 p-distance ANALYSIS

The simplest way to measure the divergence between two sequences is to count the number of different nucleotides in relation to the total length of the sequences. This measure is the p-distance. The p-distance value, however, not always reflect the true level of divergence between sequences. In fact, in the cases of multiple mutations in the same position and back-mutation (A to C to A) the p-distance detects only one change. As a consequence, the p-distance underestimates the real genetic distance, the real number of substitutions per site that occurred. In case of saturation, the phylogenetic signal is lost and the resulting phylogenetic inference is wrong, regardless of the phylogenetic reconstruction method used.

Here a graphical exploration tool is used. ML-estimated distances (*i.e.* GTR+G) (X) against uncorrected p-distances (based on observed substitutions) (Y) **(Fig. 15)**. If there is not saturation, the p-distance value is almost identical to the corrected genetic distance, and the slope of the line, R^2, is almost 1. **Fig. 15 (A)** shows the plot for the *cox1* with all its codon positions, where the R^2 value was much smaller than 1. In **(B)**, by plotting only the first and second *cox1* codon positions, without its third codon positions, the R^2 value was 0.81. In **(C)**, by plotting only the second codon positions of the *cox1* gene, the R^2 was almost 1 (0.89). **(D)** The plot of the p-distance value plotted against the ML corrected genetic distance for the 18S gene dataset showed any difference between the p-distance value and the corrected genetic distance (R^2=0.99).

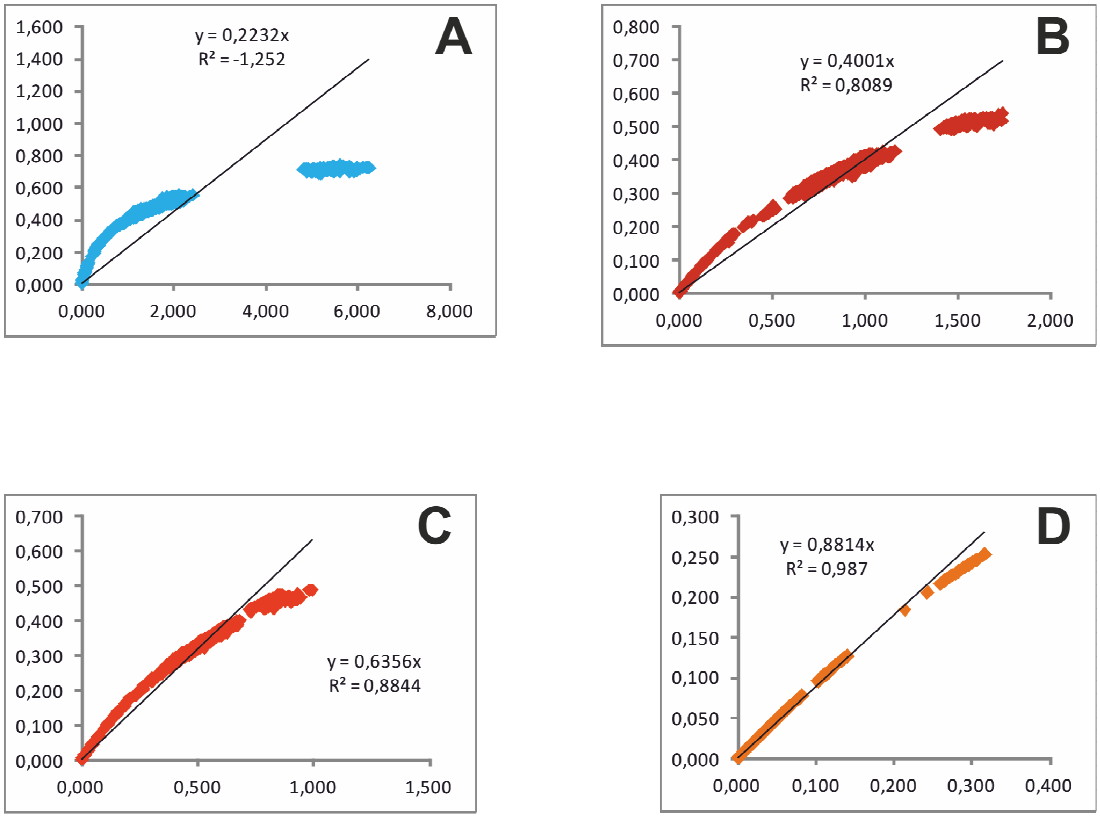


Fig. 15. ML-estimated distances (*i.e.* GTR+G) (X) against uncorrected p-distances (based on observed substitutions) (Y)**. (A)** *cox1* gene with all its codon positions. **(B)** *cox1* gene with the first and second codon positions, without the third codon positions. **(C)** *cox1* gene with only to the second codon positions. **(D)** 18S gene dataset.

5.6 PHYLOGENETIC ANALYSIS

Phylogenetic analysis was inferred by four methods: Neighbour-Joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML) and Bayesian Inference (BI). Computationally problems occurred during the BI method and so no results were obtained. MP inference showed poor resolution because it produced many equally parsimonious trees. As I reported in the previous **Paragraph 5.5**, the *cox1* gene third codon positions were largely saturated and the largest amount of the phylogenetic signal was stored within the second codon positions. Consequently, the phylogenetic analysis was inferred from the *cox1* gene, from the first and the second codon positions of the *cox1* gene without the third codon positions, from only the second codon positions of the *cox1* gene, and from the amino acidic sequences of the *cox1* gene. Phylogeny was also inferred from the 18S gene as a reference, using sequences already obtained in the laboratory where I was working.

Regarding both the *cox1* and 18S genes phylogeny, all the *Euplotes* strains were correctly portioned into species in all the three methods tested. Moreover, all the species formed monophyletic clusters with great statically support in all the inference. However, the *E. crassus* strains PLR1 and SL2 always formed a group apart with bootstrap support values of almost 100% in all the cases.

Regarding the phylogenetic analysis inferred from the *cox1* gene, all the species, with the only exception of the species *E. daidaleos* and *E. focardii*, were divided into usua groups, with great statistically significant in all cases. In fact, these values were generally of 98-100% bootstrap, and almost never below of 80% bootstrap. Given that this division into groups and the bootstrap values that support it remained more or less constant, this division of the *Euplotes* species into several groups will be described in detail only in the phylogeny inferred from the fully *cox1* gene. In the other cases, only the differences from this analysis will be described.

As regards the *cox1* phylogenetic tree inferred from all the codon positions of the gene **(Fig. 16)**, all the *Euplotes* strains were correctly portioned into species. The group comprising of the species *E. crassus, E. vannus* and *E. minuta* clustered in the apex of the phylogenetic tree. Within the upper clade of *E. crassus* it was possible to recognize three distinct groups. Group 1, strains: ST,

POR7, SAFR*4, SF8, PMEX13*1S, and Beirut3*4 (node n°1; bootstrap values: NJ = 100%, ML = 98%, MP = 100%). Group 2, strains: MAL1, Kuwait/4CCAP, and GRG4. This group was highlighted only by the NJ inference with 100% bootstrap value (node n°3; bootstrap values: NJ = 100%, ML = MP = below 50%). Group 3, strains: Shane5, GCHI-1B, OM3E3, D2, Oyster3*5, and EC4 (node n°5; bootstrap values: NJ = 100%, ML = below 50%, MP = 100%). Also within the *E. vannus* species, it was possible to distinguish two groups as the SML strain was apart from the other two conspecific strains (node n°8; bootstrap values: NJ = 100%, ML = 98%, MP = 100%). Within *E. minuta* species, the IM2 and GRG4 strains were clearly separated from the other conspecific strains (node n°13; bootstrap values: NJ = 100%, ML = 94%, MP = 100%). This was true also for the *E. minuta* strains CCAP1624/13 and Kling4 (node n°14; bootstrap values: NJ = 100%, ML = 81%, MP = 100%). The group comprising of the species *E. crassus*, *E. vannus* and *E. minuta* was closely related to the sister taxon comprising of the species *E. focardii* and *E. quinquecarinatus.* However, this relationship had a bootstrap support value greater than 50% only in the NJ inference (node n°15; bootstrap values: NJ = 74%, ML = MP = below 50%). Inside the *E. quinquecarinatus* species, it was possible to distinguish two groups, with the strains DAK5, MR13, and OMAN2 (node n°20; bootstrap values: NJ = 100%, ML = 88%, MP = 100%) apart from the other two strains GBS3 and GBS-I3. The group comprising of the species *E. focardii* and *E. quinquecarinatus* was closely related to the sister taxon comprising of the species *E. euryhalinus* and *E. gracilis*. However, the sister taxon relationship between *E. euryhalinus* and *E. gracilis* had not a bootstrap value higher than 50% in any phylogenetic inference (node n°24; bootstrap values: NJ = ML = MP = below 50%). The morphospecies *E. gracilis* was divided into two groups, the first contained the GPO4(2) strain (node n°22; bootstrap values: NJ = 100%, ML = 88%, MP = 100%), while the second contained the ATCC50191 and GNI1 strains (node n°23; bootstrap values: NJ = ML = MP = 100%). The species *E. euryhalinus* was divided in two groups too, with the strains ADC1 and WB11 clustered apart (node n°27; bootstrap values: NJ = ML = MP = 100%) from the other conspecific strains 3bILb3, 4-11Ilb5, and LAP1 (node n°26; bootstrap values: NJ = 100%, ML = 98%, MP = 99%). The group comprising of the species *E. euryhalinus* and *E. gracilis* was closely related with a sister relationship to the group comprising of the species *E. raikovi* and *E. nobilii* (node n°35; bootstrap values: NJ = 100%, ML = 98%, MP = 56%). The species *E. raikovi* and *E. nobilii* were both divided into more groups with the *E. raikovi* strain SMAaj3 divided (node n°28; bootstrap values: NJ = 100%; ML = 85%; MP = 100%) from the other *E. raikovi* strains 4MNM, Myy1, Biod2, and LPSA5 on one side (node n°30; bootstrap values: NJ = 100%;

ML = 97%; MP = 100%) and GA8 and PCE1 on the other side (node n°31; bootstrap values: NJ = 100%; ML = 98%; MP = 100%). About *E. nobilii*, the strains 3QAN7, 3QN2, 1QN1, 5QAA15, and 4Pyrm4 formed a group apart (node n°34; bootstrap values: NJ = 100%; ML = 92%; MP = 100%) from the strains 1QAA, EurhB, QAN1, 2QAN1, PNA-1h, and Sop (node n°33; bootstrap values: NJ = 100%; ML = 91%; MP = 100%). The group comprising of the species *E. raikovi* and *E. nobilii* in turn were the sister taxon of the species *E. rariseta* and *E. polaris*, which formed a well bootstrap supported sister taxon clade (node n°43; bootstrap values: NJ = 96%; ML = 98%; MP = 100%). The species *E. rariseta* and *E. polaris* were both divided into two groups. The first group of *E. rariseta* contained the strains FSl1 and Smund2 (node n°39; bootstrap values: NJ = 100%; ML = 100%; MP = 100%), while the second group contained the strains BR1 and MAM30 (node n°40; bootstrap values: NJ = ML = 100%; MP = 99%). The first group of *E. polaris* contained the strains Eup.Scon and Ila7policlonale (node n°28; bootstrap values: NJ = 100%; ML = below 50%; MP = 89%), while the second group contained the strains EdPoB02 and Split3 (node n°41; bootstrap values: NJ = ML = MP = 100%). The *E. magnicirratus* strains 3SC10 and CO were divided from the Miya3 strain (node n°45; bootstrap values: NJ = 100%; ML = 99%; MP = 100%). The group comprising of the species *E. harpa*, *E. daidaleos*, and *E. woodruffi* clustered basal to the other species. The relationship within these three morphospecies was largely bootstrap supported (node n°49; bootstrap values: NJ = 99%; ML = 92%; MP = 94%). The species *E. harpa* was clearly divided into two groups. The *E. harpa* strains GS-4, Nap3, ALM10, and ALM3 were separated (node n°47; bootstrap values: NJ = 100%; ML = below 50%; MP = 99%) from the strains Serch1, BaI5, and RB22 (node n°48; bootstrap values: NJ = 100%; ML = 85%; MP = 100%). The *E. woodruffi* strains MS-3 and SydEU6 formed a group apart from the other two conspecific strains (node n° 53; bootstrap values: NJ = 100%; ML = 98%; MP = 100%).

As regards the *cox1* phylogenetic tree inferred from only the first and the second codon positions without the third codon positions **(Fig. 17)**, all the morphospecies formed monophyletic clusters with great statically support. The group comprising of the species *E. crassus*, *E. vannus* and *E. minuta* clustered in the apex of the phylogenetic tree, and was closely related to the sister taxon comprising of the species *E. rariseta* and *E. gracilis*. However, the sister taxon relationships between *E. rariseta* and *E. gracilis* had not bootstrap values higher than 50% in any phylogenetic reconstruction (node n°15; bootstrap values: NJ = ML =MP = below 50%). The group comprising of the species *E. rariseta* and *E. gracilis* was closely related to the sister taxon comprising of the

species *E. magnicirratus* and *E. euryhalinus*. These species, *E. magnicirratus* and *E. euryhalinus*, were closely related to the sister taxon comprising of the three species *E. harpa*, *E. daidaleos*, and *E. woodruffi*. These three species were closely related to *E. polaris*, but with no significant bootstrap support values (node n°37; bootstrap values: NJ = ML = MP = below 50%). *E. raikovi*, *E. nobilii*, *E. quinquecarinatus*, and *E. focardii* clustered basal to the other species.

As regards the *cox1* phylogenetic tree inferred from only its second codon positions **(Fig. 18)**, all the *Euplotes* strains were correctly portioned into species. Moreover, all the morphospecies formed monophyletic clusters with great statically support. The sister taxon comprising of the species *E. crassus*, *E. vannus* and *E. minuta* clustered in the apex of the phylogenetic tree. Within the species *E. minuta*, the usual division of the strains was not confirmed, as the strain MAR11 was grouped with the conspecific strains IM2 and GAR4. The sister taxon comprising of the species *E. crassus*, *E. vannus* and *E. minuta* was closely related to the species *E. focardii* and *E. gracilis*. The species *E. gracilis* was closely related to the sister taxon comprising of the species *E. magnicirratus* and *E. euryhalinus.* This sister taxon relatioships was not well bootstrap supported (node n°22; bootstrap values: NJ = ML = MP = below 50%). The sister taxon comprising of the species *E. magnicirratus* and *E. euryhalinus* was close to *E. raikovi*. *E. raikovi* was closely related to the species *E. rariseta* in turn. The species *E. rariseta* was closely related to the sister taxon comprising of the species *E. harpa*, *E. daidaleos*, and *E. woodruffi*. This sister taxon relationships was well bootstrap supported (node n°35; bootstrap values: NJ = 99%; ML = MP = 95%). These three sister taxon species were closely related to *E. polaris*, without a significant bootstrap consensus (node n°38; bootstrap values: NJ = ML = MP = below 50%). *E. polaris* this time was not divided into two groups. *E. nobilii* and *E. quinquecarinatus* clustered basal to the other species.

As regards the *cox1* phylogenetic tree inferred from its amino acidic sequences **(Fig. 19)**, all the *Euplotes* strains were correctly portioned into species. Moreover, all the morphospecies formed monophyletic clusters with great statically support. The sister taxon comprising of the species *E. crassus*, *E. vannus* and *E. minuta* clustered in the apex of the phylogenetic tree and was closely related to *E. focardii* and the sister taxon comprising of the species *E. gracilis* and *E. euryhalinus*. The species *E. gracilis* and *E. euryhalinus* were closely related to the species *E. magnicirratus* and *E. rariseta*. The *E. rariseta* species was closely related to the sister taxon comprising of the species

E. harpa, *E. daidaleos*, and *E. woodruffi*. This sister taxon relationships was well bootstrap supported (node; n°30 bootstrap values: NJ = 92%; ML = 82%; MP = 72%). These three sister taxon species were closely related to *E. polaris*, who was not divided into two groups. The species *E. raikovi*, *E. nobilii* and *E. quinquecarinatus* clustered basal to the other species.

As regards the 18S gene phylogenetic tree **(Fig. 20)**, all the *Euplotes* strains were correctly portioned into species. Moreover, all the morphospecies formed monophyletic clusters with great statically support. The sister taxon comprising of the species *E. crassus*, *E. vannus* and *E. minuta* clustered in the apex of the phylogenetic tree. Within the upper clade of *E. crassus* it was possible to recognize only two distinct groups. In fact the *E. crassus* strains GCHI-1B, GRG4, Kuwait/4CCAP, and MAL1 formed a group apart (node n° 3; bootstrap values: NJ = 98%; ML = 99%; MP = 97%) from the conspecific strains (node n° 1; bootstrap values: NJ = 65%; ML = 88%; MP = below 50%). These sister taxon comprising of the species *E. crassus*, *E. vannus* and *E. minuta* was closely related to the sister taxon comprising of the species *E. focardii* and *E. quinquecarinatus*. *E. focardii* and *E. quinquecarinatus* were closely related to the sister taxon formed by the species *E. rariseta* and *E. gracilis*. However, the sister taxon relationships between *E. rariseta* and *E. gracilis* had not statically relevance (node n°15; bootstrap values: NJ = ML = MP = below 50%). *E. rariseta* and *E. gracilis* were closely related to the sister taxon comprising of the species *E. magnicirratus* and *E. euryhalinus*, whose sister taxon relationships was well bootstrap supported (node n°19; bootstrap values: NJ = 68%; ML = 100%; MP = 84%). *E. magnicirratus* and *E. euryhalinus* were closely related to sister taxon comprising of the three species *E. harpa*, *E. daidaleos*, and *E. woodruffi* (node n°23; bootstrap values: NJ = ML = 100%; MP = 99%). These three species were closely related to the sister taxon comprising of the species *E. raikovi* and *E. nobilii* (node n°29; bootstrap values: NJ = ML = MP = 100%). *E. Polaris* clustered basal to the other species (node n°33; bootstrap values: NJ = ML = 100%; MP = 96%).

I inferred the phylogeny from the two *Euplotes cox1* gene sequences available on GenBank together with all the *E. crassus* and the *E. minuta* strains that I tested in this work. Obviously, I expected that the *E. crassus* from GenBank would cluster within the *E. crassus* strains of my dataset. Similarly, I also expected that the *E. minuta* from GenBank would cluster within the *E. minuta* strains of my dataset. As is shown in **(Fig. 21)**, my hypothesis was not confirmed. In fact, both the *E. crassus* and the *E. minuta* available on GenBank clustered within the *E. crassus* strains I analyzed in this study, with bootstrap support values of 100% in both cases. The *cox1 E. crassus* sequence on GenBank (GQ903131.1) clustered within the *E. crassus* group 1. The *cox1 E. minuta* sequence on GenBank (GQ903130.1) clustered within the *E. crassus* group 3.

	NJ	ML	MP		NJ	ML	MP
1	100	98	100	29	100	89	100
2	95	--	--	30	100	97	100
3	100	--	--	31	100	98	100
4	100	84	100	32	100	92	100
5	100	--	100	33	100	91	100
6	100	--	--	34	100	92	100
7	68	--	--	35	100	98	56
8	100	98	100	36	--	--	--
9	100	100	100	37	--	--	--
10	84	76	68	38	92	99	--
11	99	99	97	39	100	100	100
12	100	99	100	40	100	100	99
13	100	94	100	41	100	100	100
14	100	81	99	42	100	--	89
15	74	100	100	43	96	98	100
16	100	100	100	44	96	98	100
17	57	--	--	45	100	99	100
18	100	100	99	46	100	100	100
19	100	100	--	47	100	--	99
20	100	88	100	48	100	85	100
21	--	--	--	49	99	92	94
22	100	88	100	50	100	93	97
23	100	100	100	51	100	98	100
24	--	--	--	52	100	96	100
25	100	75	97	53	100	98	100
26	100	99	100	54	100	100	100
27	100	100	100	55	100	100	100
28	100	85	100				

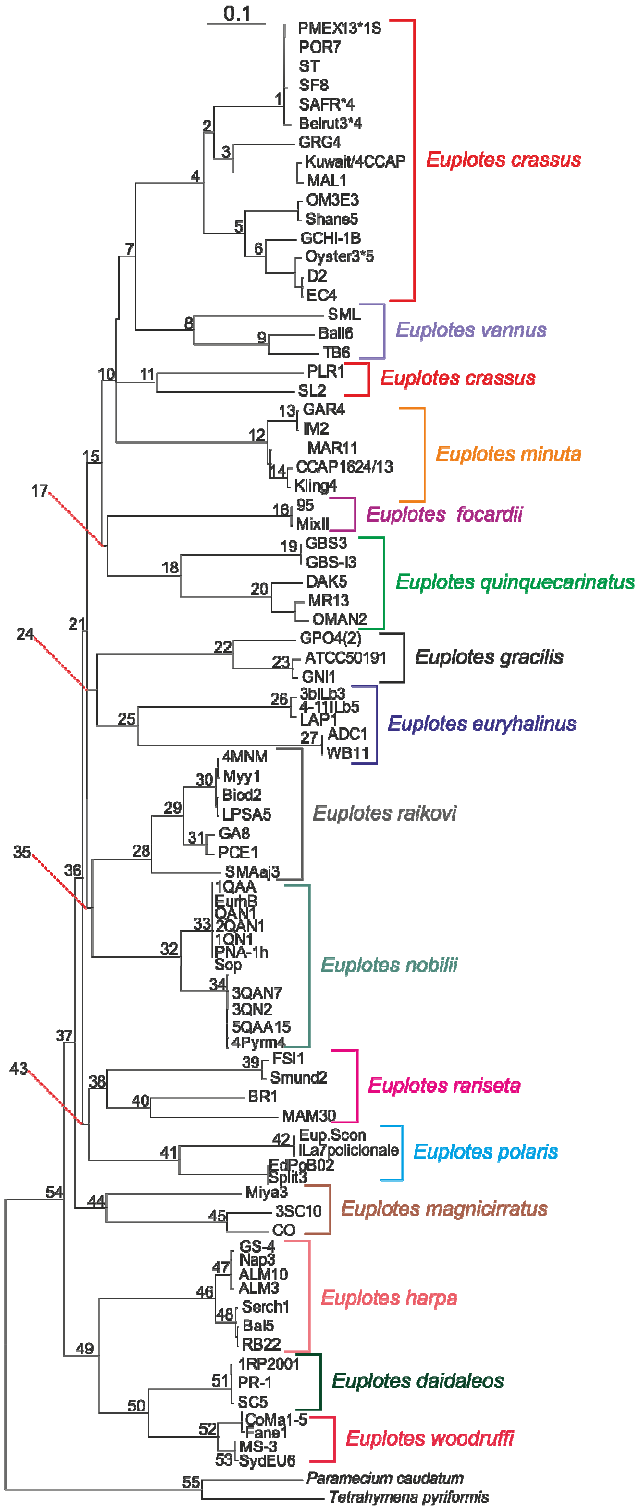


Fig. 16. Phylogenetic tree inferred from the *cox1* gene. The bar indicates 10 nucleotide substitutions per 100 nucleotides. Nodes are numbered progressively starting from the apex of the phylogenetic tree.The numbers at the nodes are bootstrap percentages from 1000 replicates of NJ, ML and MP inference respectively. They were reported in the table showed on the left. Only bootstrap values higher than 50% were shown. The *cox1* sequences of the outgroups species *Paramecium caudatum* and *Tetrahymena pyriformis* were taken from the GenBank/EMBL databases (NC_014262.1 and NC_000862.1, respectively).

	NJ	ML	MP		NJ	ML	MP
1	100	89	89	28	100	100	99
2	94	--	--	29	58	--	--
3	100	57	51	30	99	98	98
4	100	75	99	31	100	100	99
5	100	66	61	32	100	--	--
6	89	--	69	33	100	99	99
7	100	92	99	34	100	100	100
8	100	99	99	35	100	100	99
9	50	--	55	36	100	100	99
10	100	91	99	37	--	--	--
11	89	--	89	38	--	--	--
12	100	93	99	39	--	--	--
13	100	97	99	40	100	100	99
14	100	89	96	41	99	99	94
15	--	--	--	42	100	97	61
16	79	99	77	43	100	86	99
17	100	100	99	44	100	97	99
18	100	100	99	45	--	--	--
19	100	100	99	46	100	89	89
20	100	99	99	47	99	64	64
21	--	--	--	48	100	94	94
22	--	--	--	49	100	100	99
23	--	--	--	50	100	100	99
24	99	100	94	51	100	100	100
25	100	98	99	52	100	100	100
26	100	100	99	53	100	100	99
27	100	94	99	54			

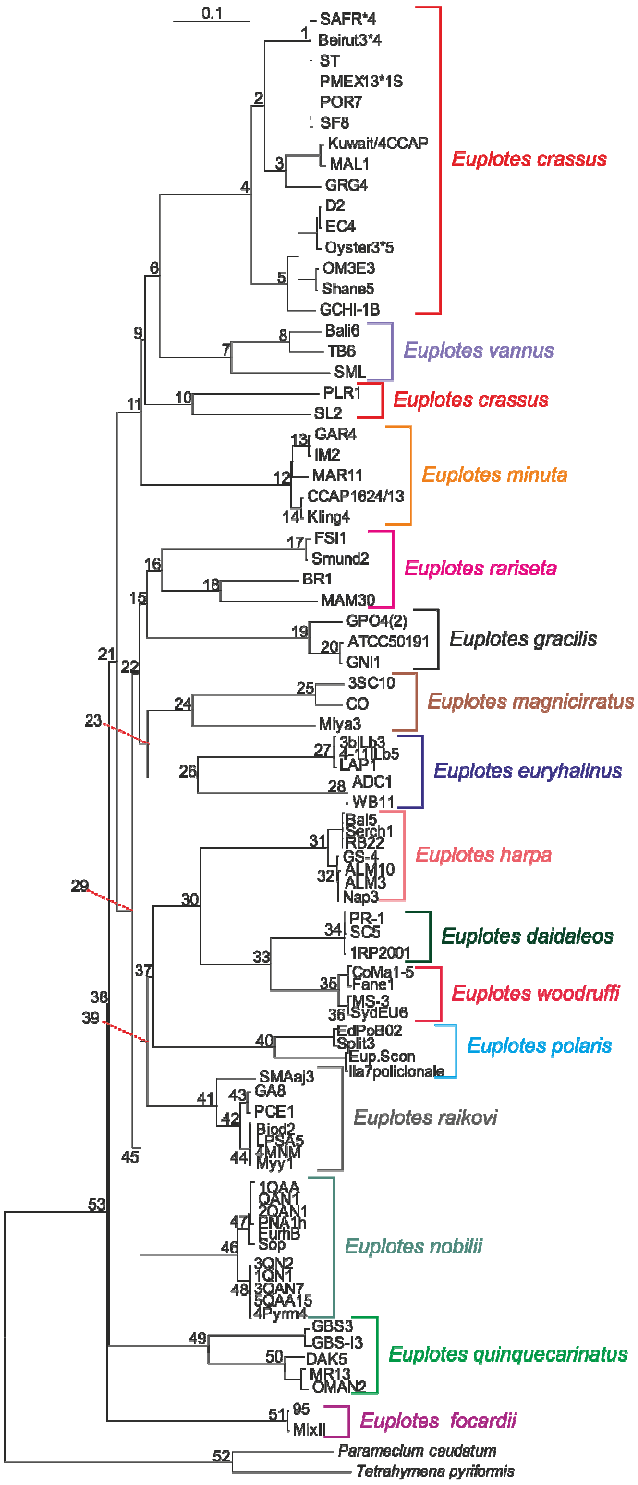


Fig. 17. Phylogenetic tree inferred from the first and second codon positions of the *cox1* gene, without the third codon positions. The bar indicates 10 nucleotide substitutions per 100 nucleotides. Nodes are numbered progressively starting from the apex of the phylogenetic tree.The numbers at the nodes are bootstrap percentages from 1000 replicates of NJ, ML and MP inference respectively. They were reported in the table showed on the left. Only bootstrap values higher than 50% were shown. The *cox1* sequences of the outgroups species *Paramecium caudatum* and *Tetrahymena pyriformis* were taken from the GenBank/EMBL databases (NC_014262.1 and NC_000862.1, respectively).

	NJ	ML	MP		NJ	ML	MP
1	100	96	99	26	100	94	99
2	76	--	--	27	99	96	65
3	99	54	--	28	96	62	96
4	100	98	98	29	99	57	94
5	99	96	98	30	72	--	--
6	73	69	--	31	100	--	78
7	100	99	98	32	100	98	99
8	100	99	97	33	--	100	97
9	--	75	54	34	99	--	--
10	100	100	99	35	99	95	95
11	70	--	74	36	100	99	99
12	100	98	99	37	100	--	--
13	100	66	56	38	--	--	--
14	--	74	--	39	--	--	--
15	100	100	99	40	100	94	99
16	100	100	99	41	100	99	99
17	100	98	93	42	100	97	99
18	--	--	--	43	100	99	99
19	98	95	96	44	100	94	99
20	100	95	99	45	100	85	100
21	--	--	--	46	100	94	99
22	--	--	--	47	87	97	52
23	--	--	--	48	67	100	--
24	99	100	98	49	100	100	99
25	100	100	99	50	100	100	--

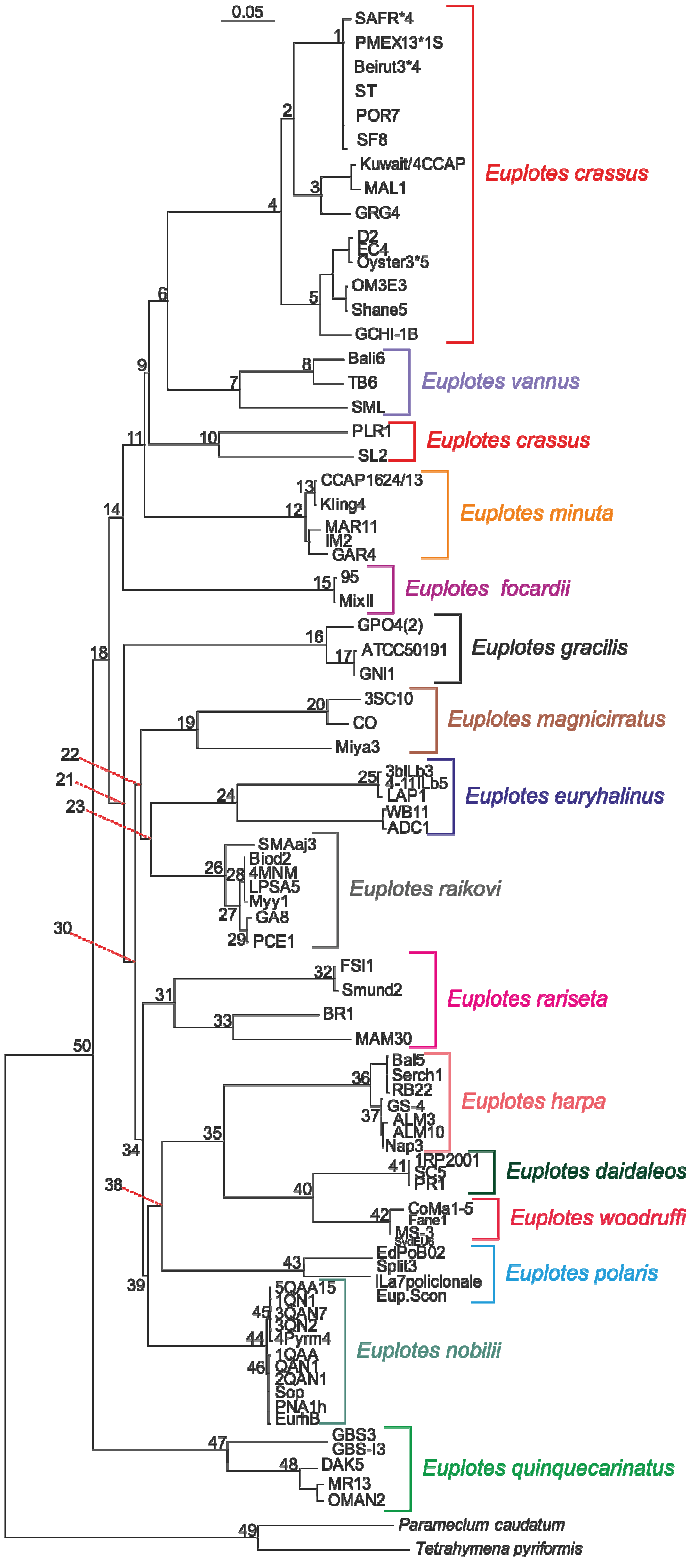


Fig. 18. Phylogenetic tree inferred from the second codon positions of the *cox1* gene. The bar indicates 5 nucleotide substitutions per 100 nucleotides. Nodes are numbered progressively starting from the apex of the phylogenetic tree.The numbers at the nodes are bootstrap percentages from 1000 replicates of NJ, ML and MP inference respectively. They were reported in the table showed on the left. Only bootstrap values higher than 50% were shown. The *cox1* sequences of the outgroups species *Paramecium caudatum* and *Tetrahymena pyriformis* were taken from the GenBank/EMBL databases (NC_014262.1 and NC_000862.1, respectively).

	NJ	ML	MP		NJ	ML	MP
1	88	99	100	27	100	89	100
2	100	98	99	28	100	100	99
3	100	50	--	29	--	--	--
4	97	78	--	30	92	82	72
5	100	65	64	31	98	73	82
6	61	69	--	32	100	--	--
7	100	99	99	33	100	84	90
8	100	96	99	34	100	87	99
9	91	90	86	35	100	98	99
10	100	97	93	36	100	92	96
11	50	--	--	37	100	100	100
12	100	98	100	38	--	--	--
13	100	90	86	39	--	--	--
14	100	80	--	40	--	--	--
15	100	100	99	41	100	100	100
16	--	--	--	42	100	95	99
17	--	--	--	43	100	95	99
18	100	93	100	44	97	96	89
19	100	93	92	45	99	98	98
20	100	83	98	46	--	--	--
21	100	88	99	47	100	100	99
22	100	100	100	48	98	98	98
23	--	--	--	49	99	100	99
24	100	100	99	50	100	98	100
25	100	100	100	51	100	97	98
26	96	--	50	52	100	100	100

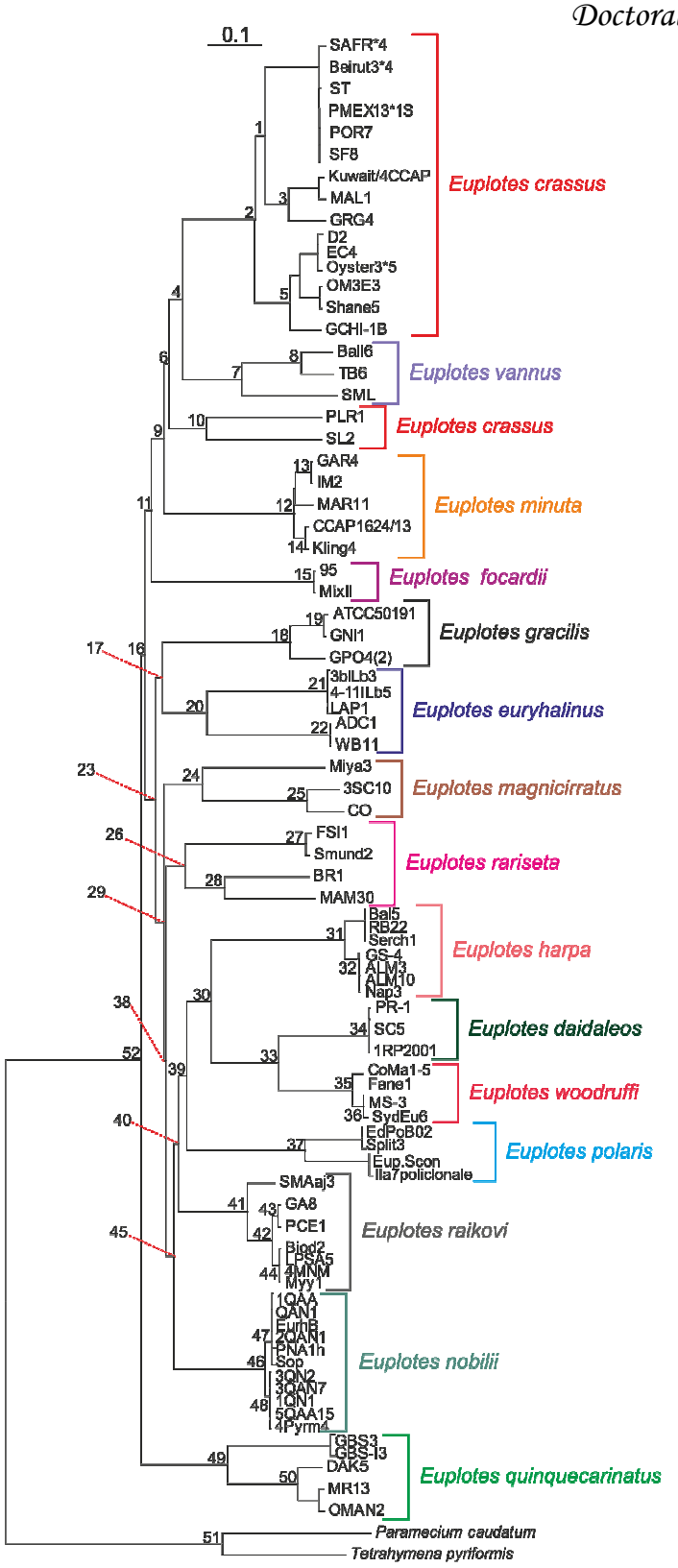


Fig. 19. Phylogenetic tree inferred from the amino acidic sequences of the cox1 gene. The bar indicates 10 nucleotide substitutions per 100 nucleotides. Nodes are numbered progressively starting from the apex of the phylogenetic tree.The numbers at the nodes are bootstrap percentages from 1000 replicates of NJ, ML and MP inference respectively. They were reported in the table showed on the left. Only bootstrap values higher than 50% were shown. The cox1 sequences of the outgroups species Paramecium caudatum and Tetrahymena pyriformis were taken from the GenBank/EMBL databases (NC_014262.1 and NC_000862.1, respectively).

	NJ	ML	MP		NJ	ML	MP
1	65	88	--	18	--	--	--
2	99	85	--	19	68	100	84
3	98	99	97	20	100	100	99
4	86	100	98	21	100	100	99
5	85	90	89	22	100	100	100
6	100	100	95	23	100	100	99
7	71	89	98	24	100	100	100
8	99	100	98	25	100	100	100
9	99	89	--	26	100	100	100
10	100	100	99	27	100	100	100
11	100	100	99	28	100	100	100
12	100	100	99	29	100	100	100
13	100	100	--	30	100	100	100
14	70	79	79	31	100	100	100
15	--	--	--	32	100	100	100
16	100	100	99	33	100	100	96
17	100	100	100	34	100	100	100

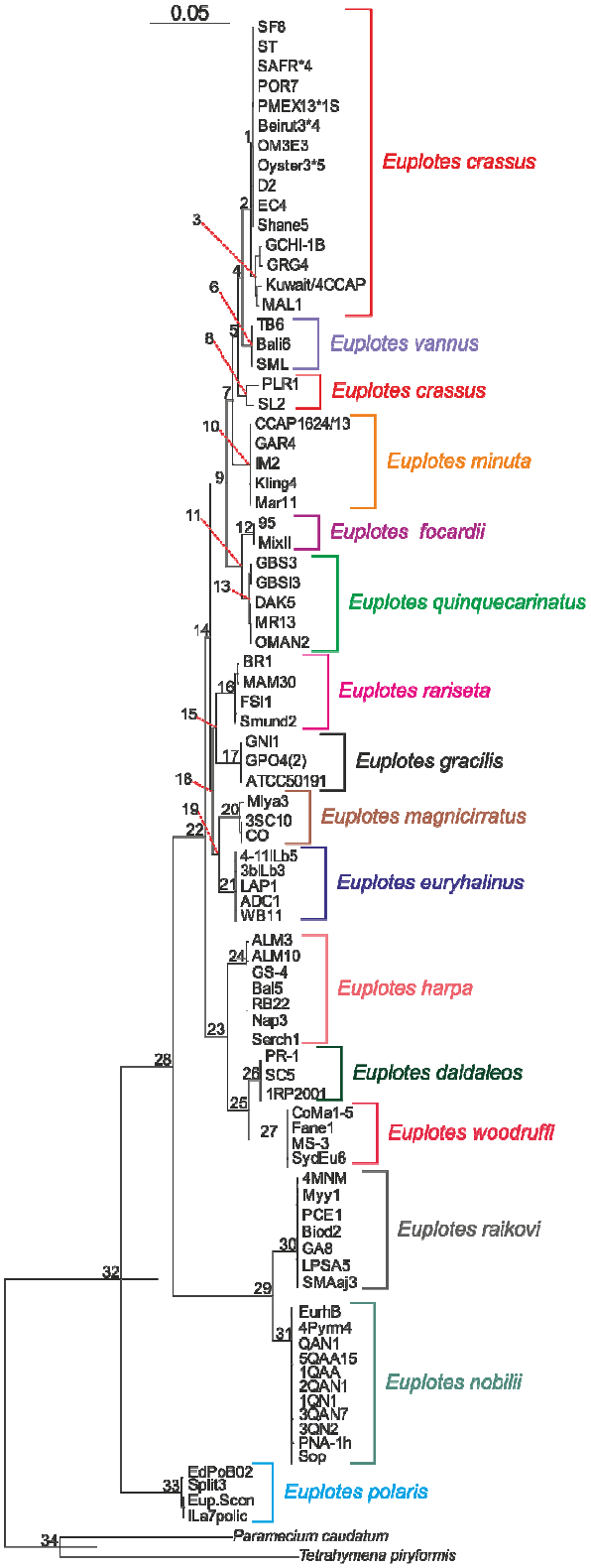


Fig. 20. Phylogenetic tree inferred from the 18S gene. The bar indicates that 5 substitution per 100 nucleotidic sites occurred. Nodes are numbered progressively starting from the apex of the phylogenetic tree.The numbers at the nodes are bootstrap percentages from 1000 replicates of NJ, ML and MP inference respectively. They were reported in the table showed on the left. Only bootstrap values higher than 50% were shown. The 18S sequences of the outgroups species *Paramecium caudatum* and *Tetrahymena pyriformis* were taken from the GenBank/EMBL databases (JF304170.1 and EF070254.1, respectively).

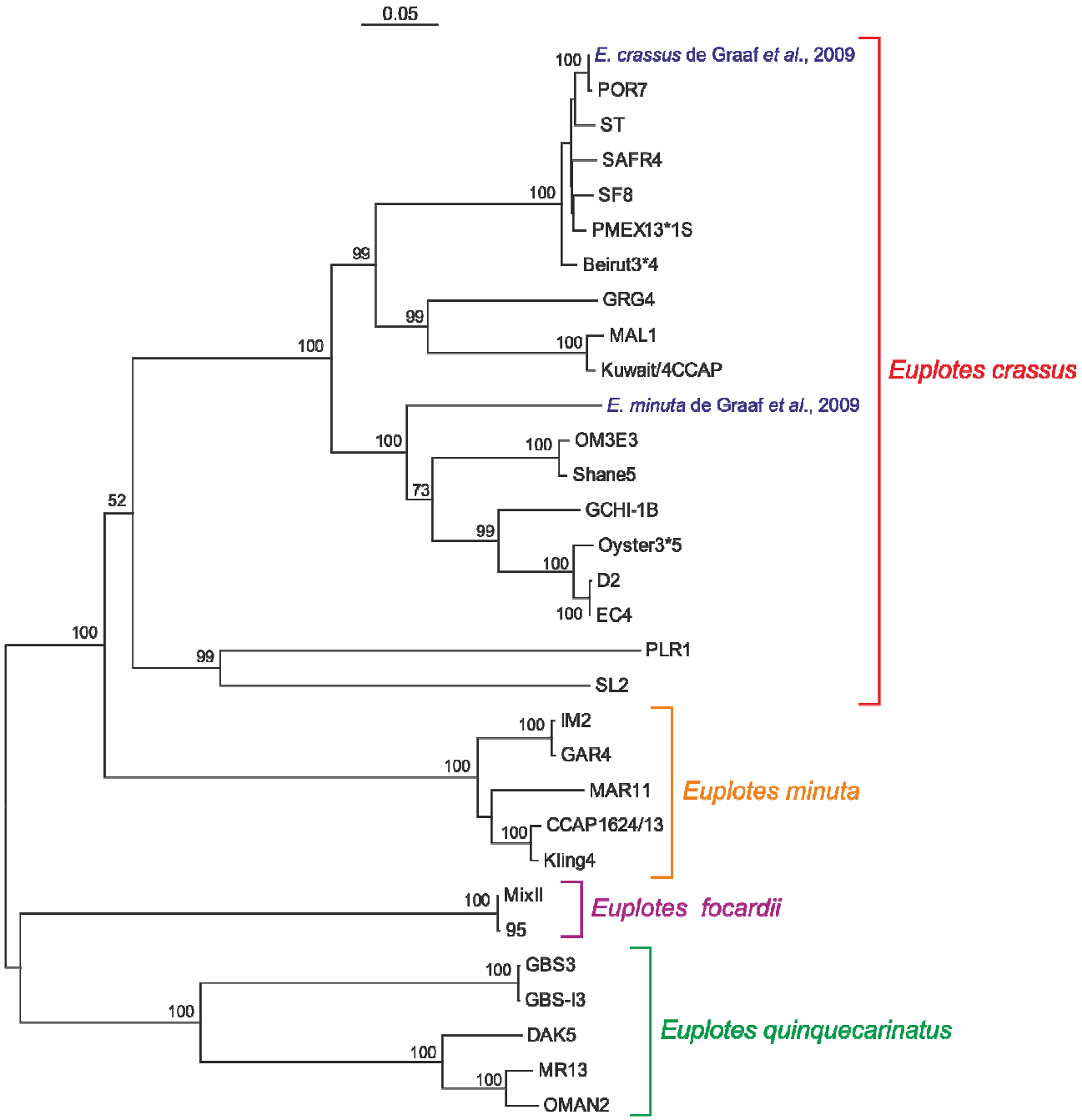


Fig. 21. Phylogenetic tree inferred from the *cox1* nucleotide sequences of *E. crassus* (GQ903131.1) and *E. minuta* (GQ903130.1) strains available on GenBank/databases, compared with the homologous sequences of *E. crassus* and *E. minuta* strains of my dataset. The bar indicates 5 nucleotide substitutions per 100 nucleotides. The numbers at the node are bootstrap percentages from 1000 replicates of NJ. The *cox1* sequences of the outgroups species come from my dataset.

6 DISCUSSION

6.1 SEQUENCES ANALYSIS OF THE *cox1* GENE

I obtained *cox1* amplified fragments of differing lengths depending on the tested species **(Table 6)**. Even within some *Euplotes* species I obtained amplified fragments of differing lengths, in accordance with de Graaf and collaborators (2009) [40], who sequenced the mitochondrion genomes of two *Euplotes* species, *E. crassus* and "*E. minuta*". My work, in fact, demonstrated that the *E. minuta* on GenBank is indeed *E. crassus* **(Paragraph 6.2)**. Also the species identified as *E. crassus* and *E. minuta* on GenBank showed *cox1* genes of different lengths: 2945 bp and 3611 bp, respectively. These two *Euplotes* species in fact, did not differ from each other in term of gene order, but they differed in terms of gene size.

The *cox1* gene nucleotidic sequences analyzed in this study **(Table 8)** had a lower G+C content (33.8%) than other ciliated protists, such as those belonging to the genus *Tetrahymena* (42.9%) [126]. However, this result was in accordance with mean nucleotidic frequencies of the *E. crassus* and the so identified *E. minuta* on GenBank/databases, whose G+C contents were 34.7% and 36.0%, respectively [40]. It is well known that higher G+C content is associated with coding DNA sequences and advanced level of genomic organization. When other *Euplotes* mitochondrial genome will be available, other speculations can be made regarding the evolutionary history of ciliated protists.

The percentages of the *cox1* nucleotidic sequence divergence between and within the *Euplotes* species were calculated using the K2P distance model of evolution **(Table 9)**. Values thus obtained were compared to those in literature both in animal organisms and protists. The percentage of *cox1* nucleotidic sequence divergence between species was 62.7%, which is so far the higher value ever found in literature. In fact, in general, interspecific *cox1* sequence divergence values of various groups of animals were about 10%, while intraspecific sequence divergences were typically below 1%. For example, percentages of *cox1* interspecific and intraspecific sequence variation were respectively 7.93% and 0.43% in North American birds [13], 9.93% and 0.39% in Australian fishes [21]. The DNA barcoding of *Tetrahymena* [126] confirmed these results and showed 11.3% of interspecific *cox1* sequence variation and <0.65% of intraspecific *cox1* sequence variation. The DNA barcoding gap, defined as the division between the percentage of interspecific and intraspecific *cox1* sequence variation, generally assumes a fixed value comprised between 10 and

20 in almost all DNA barcoding studies **(Table 10)**. Otherwise, in the *Euplotes* tested species, this DNA barcoding gap law was not respected. Consequently, the DNA barcoding gap did not assume a fixed value, and it ranged from 1.4 to 125.4. Moreover, interspecific and intraspecific *cox1* sequence divergence values never overlapped and were always well separated. By treating the groups within the tested *Euplotes* species they were divided by the phylogenetic analysis as independent evolutionary units **(Paragraph 5.6)**, the range of the DNA barcoding gap values decreased, but it did not become consistent to those in literature. In fact, it ranged from 1.2 in the *E. crassus* strains PLR1 and SL2 to 52.2 in the group formed by the *E. euryhalinus* strains 3bILb3, 4-11ILb5 and LAP1. This means the species cannot be distinguished on the basis of the threshold method. However, there was not always any overlap between the intraspecific and the interspecific *cox1* variation. The intraspecific *cox1* sequence variation within the *E. crassus* strains PLR1 and SL2 still remained huge (49.8%).

6.2 PHYLOGENETIC ANALYSIS

Overall, the Likelihood mapping and the p-distance analysis all together indicated that only the *cox1* third codon positions were widely saturated. That allowed me to exclude them from the phylogenetic inference. The *cox1 s*econd codon positions contained most of the phylogenetic information. Any saturation affected the 18S gene **(Fig. 14** and **Fig. 15)**.

The *cox1* was both a really fast evolving gene and protein. The *cox1* rate of sequence evolution decreased when only the second codon positions were analyzed. This is normal, in fact the saturation analysis showed that most substitutions occurred on the third codons positions, while the *cox1* second codon position were the most stable.

Surprisingly, the Bayesian Inference did not work at all. Since in literature there was a great enthusiasm on the Bayesian methods for estimating tree [86], I expected it would be the method that would give me the result that best fitted the data. Probably, the saturation that affected the *cox1* gene third codon positions and the large level of heterogeneity that characterized the *cox1* gene dataset meant that no results were obtained. This would also explain why the Maximum Parsimony inference provided a huge number of equally parsimonious trees and was not able to find the optimal tree. Although the Neighbor-Joining analysis and the K2P model have been widely criticized, I have used them for comparison to other ciliated protists DNA barcoding research [126] **(Table 2)**. From a "DNA barcoding point of view" rather than a "phylogenetic point of view", the NJ analysis is most appropriate because it provided consistent results in seconds, while the other methods took hours or even days, depending on the dataset sizes. Consequently, if the goal is to do the DNA barcoding, NJ is the better analysis. It would not make sense to use algorithmic methods that employ times as long as the morphological approach to identify a species.

All the species were monophyletic with almost always 100% bootstrap support values in the *cox1* phylogenetic inference. The phylogenetic inference from the 18S gene in this study was in accordance with the 18S gene phylogeny of *Euplotes* in literature [141, 156, 157]. In many cases, the *cox1* phylogeny provided alternative topology. However, those were not statistically significant. In the phylogeny inferred from all the codon positions of the *cox1* gene, any deep nodes were bootstrap supported. Of course it was a consequence of the saturation that occurred in the third

codon positions of the *cox1* gene. In the phylogenetic tree inferred from the *cox1* gene without its third codon positions in fact, a recovery of the 18S topology is observed. However, that topology was better statistically supported in the 18S gene phylogeny than in the *cox1* gene phylogeny. This did not happen in the phylogenetic tree inferred from only the second codon positions of the *cox1* gene.

In the *cox1* gene phylogeny all the species, with the exceptions of *E. focardii* and *E. daidaleos,* were always divided into more groups. This splitting into several groups was consistent with other DNA barcoding studies in protists [112, 126]. Almost always, in fact, DNA barcoding studies revealed a large amount of cryptic diversity.

Overall, the species *E. crassus*, *E. vannus* and *E. minuta* formed the apex clustered clade in all the phylogenetic inference. The species *E. focardii* and *E. quinquecarinatus* formed a group, which was not well stable. In fact, this relationship was confirmed only by the 18S and by the fully *cox1* genes. In the 18S phylogeny, the *E. rariseta* and *E. gracilis* species formed a sister taxon clade. This clade in the *cox1* phylogeny was confirmed only in the without *cox1* third codon positions inference. The species *E. magnicirratus* and *E. euryhalinus* formed a relatively stable sister taxon clade. This relationship was not confirmed in the phylogeny inferred from fully *cox1* codon positions and *cox1* amino acidic sequences. The species *E. harpa*, *E. daidaleos,* and *E. woodruffi* formed the second well stable clade in this phylogenetic analysis. Their sister taxon relationship, in fact, is confirmed in the whole phylogenetic inference. Also the species *E. raikovi* and *E. nobilii* formed a stable clade. The only exception was the phylogeny inferred by only the second codon positions of the *cox1* gene. The species *E. polaris* clustered basal alone to the other species only in the 18S phylogeny. Its sister taxon relationships highlighted by the *cox1* inferences were never statistically significant.

It is very interesting to point out that the *E. crassus* strains PLR1 and SL2 were always (*cox1* and 18S both) clustered apart from the other *E. crassus* strains, and they were always closer to the *E. vannus* and *E. minuta* species. Moreover, their amplified fragments by the PCR were differing in length from each other, and the intraspecific *cox1* sequence variation within them was huge (49.8%). Despite the fact that their morphological features overlapped with those of the morphospecies *E. crassus,* each of the two strains most likely represents a different unit of

evolution. This hypothesis is also confirmed by the production profile of secondary metabolites (euplotins) produced by either of these two strains (SL2). In fact, it was observed that this strain did not produce any type of euplotins, unlike the *E. crassus* species [3]. To state that the strains in question belong to real evolutionary units, additional tests must be performed, such as experiments based on the biological concept of species (breeding tests).

Probably this occurred also in the morphospecies *E. rariseta*. Regarding the strains BR1 and MAM30, which always formed a group apart in phylogenetic inference, they showed a high percentage of intraspecific nucleotidic *cox1* sequence divergence (38.9%), and their corresponding amplified fragments were differing in length. Like *E. crassus*, also this morphospecies showed intra-morphospecific variability in its secondary metabolites production that was reflected in the phylogenetic partition.

Despite the fact that *E. minuta* and *E. woodruffi* were divided into more groups by all the phylogenetic trees, they didn't show high *cox1* intraspecific sequence divergence values. Regarding *E. minuta*, any nucleotidic divergence was found between the GAR4 and IM2 strains on one side and between the Kling4 and CCAP1624/13 on the other side. This was true also for the *E. woodruffi* group formed by the MS-3 and SydEU6 strains. Probably, in these species, the number of mutations that separated the various groups is not comparable with that which separated groups in the others morphospecies.

The species *E. minuta*, whose mitochondrion genome has been made recently available on GenBank/databases, was certainly misdiagnosed and is actually *E. crassus*. Our collection of living strains of *Euplotes* is the biggest in the world. This huge collection contains various strains of *E. minuta*, which have been identified as such using various methodological approaches (morphology, morphometry, mating types, genetic, and biochemistry). So, I am absolutely sure of the correct identification of the *E. minuta* strains in this collection. All the PCR primers I designed based on the sequence of the *E. minuta* available on GenBank did not work for any *E. minuta* in collection. For that reason, I quickly understood there was an error, and that the strain of *E. minuta* available on GenBank was not actually *E. minuta*. The problem was to understand which strain it had been confused with. The answer was the application of the DNA barcoding methodology. The result was greatly bootstrap supported (100%) **(Fig. 21)**, so there is no doubt that the strain on

GenBank misdiagnosed as *E. minuta* is actually *E. crassus*. To be honest, I did not expect the strain to be *E. crassus*, because *E. minuta* is much smaller, and I was surprised by such an obvious error. It is really interesting because the Authors did not realize that instead of sequencing the mitochondrion genomes of two different species, they had sequenced the mitochondrion genomes of two strains belonging to the same morphospecies, but at two different evolutionary pathways, as it was also demonstrated by this work, by the 18S gene phylogeny, and by secondary metabolites studies. Evidently, the nucleotidic sequences and the genome organization between strains belonging to the same morphospecies were so different they did not realize the error. This is further proof that problem of identification of the *Euplotes* species is still current.

7 CONCLUSIONS

The DNA barcoding tool has been shown to be effective and thus can be used as an invaluable tool for protistologists in the future. With this work I have laid the groundwork for developing a fast, safe and inexpensive tool to identify the species of *Euplotes*. At this point one forward and one reverse primer must be designed within the amplified fragment by the PCR primers which allow direct sequencing in one single run. The DNA barcoding tool will allow us to study more deeply and discover new aspects of these fascinating free-living protists. In particular, it will speed up the discovery of cryptic species and it will lead us to have more precise information on their biogeography.

I am facing the biggest paradox in the DNA barcoding literature. The data clearly showed the great potential of the *cox1* gene as barcode for the *Euplotes* genus. Although it was not possible to define a universal threshold for the *Euplotes* genus, the interspecific and intraspecific *cox1* sequence variation were clearly separated and never overlapped. High bootstrap values of almost always 100% and never less than 98% in all the three tested phylogenetic methods strongly supported the monophyly of all the *Euplotes* morphospecies under study. Although the *cox1* gene was clearly a good barcode for the *Euplotes* genus, it was not also a good phylogenetic marker. This was mainly due to saturation that largely affected the *cox1* gene third codon positions.

Overall, the *cox1* gene strongly suggested the presence of a large amount of cryptic diversity within the all the tested *Euplotes* species, except for *E. daidaleos* and *E. focardii*. However, this was certainly due to numerical under sampling. The finding of cryptic diversity was in accordance with other barcoding studies in both animal and protists [98, 112, 126]. The amplification by PCR of fragments of differing length depending on the tested species, the high degree of sequence divergence both nucleotidic and amino acidic that occurred in the most in particularly species, and the precise division of these species in particular into more groups by all the phylogenetic trees inferred from the *cox1* gene, all together suggested the occurrence of cryptic diversity inside these species.

The species *E. minuta* whose mitochondrion genome has been made recently available on GenBank was certainly misdiagnosed and is actually *E. crassus*. Evidently, the ability to distinguish between the *Euplotes vannus-crassus-minuta* complex continues to be contentious.

8 LITERATURE CITED

1 Cervia D, Garcia-Gil M, Simonetti E, Di Giuseppe G, Guella G, Bagnoli P, Dini F: Molecular mechanisms of euplotin c-induced apoptosis: Involvement of mitochondrial dysfunction, oxidative stress and proteases. Apoptosis 2007;12:1349-1363.

2 Cervia D, Martini D, Garcia-Gil M, Di Giuseppe G, Guella G, Dini F, Bagnoli P: Cytotoxic effects and apoptotic signalling mechanisms of the sesquiterpenoid euplotin c, a secondary metabolite of the marine ciliate euplotes crassus, in tumour cells. Apoptosis 2006;11:829-843.

3 Guella G, Skropeta D, Di Giuseppe G, Dini F: Structures, biological activities and phylogenetic relationships of terpenoids from marine ciliates of the genus euplotes. Mar Drugs 2010;8:2080-2116.

4 Gotelli NJ: A taxonomic wish-list for community ecology. Philos Trans R Soc Lond B Biol Sci 2004;359:585-597.

5 Hammond P (ed Species inventory. In global biodiversity:Status of the earth's living resources, London: Chapman & Hall, 1992, pp 17-39.

6 Hawksworth DL, Kalin-Arroyo MT (eds): Magnitude and distribution of biodiversity., Cambridge University Press, 1995, pp 107-191.

7 Wagele H, Klussmann-Kolb A, Kuhlmann M, Haszprunar G, Lindberg D, Koch A, Wagele JW: The taxonomist - an endangered race. A practical proposal for its survival. Front Zool 2011;8:25.

8 Wheeler QD, Raven PH, Wilson EO: Taxonomy: Impediment or expedient? Science 2004;303:285.

9 Hebert PD, Cywinska A, Ball SL, deWaard JR: Biological identifications through DNA barcodes. Proc Biol Sci 2003;270:313-321.

10 Wilson EO: The encyclopedia of life. TRENDS in Ecology and Evolution 2003;18:77-81.

11 Tautz D: A plea for DNA taxonomy. TRENDS in Ecology and Evolution 2003;18:70-75.

12 Arnot DE, Roper C, Bayoumi RA: Digital codes from hypervariable tandemly repeated DNA sequences in the plasmodium falciparum circumsporozoite gene can genetically barcode isolates. Mol Biochem Parasitol 1993;61:15-24.

13 Hebert PDN: Identification of birds through DNA barcodes. PLoS BIOLOGY 2004;2:312-319.

14 Bisby FA, Shimura J, Ruggiero M, Edwards J, Haeuser C: Taxonomy, at the click of a mouse. Nature 2002;418:367.

15 Godfray HC, Lewis T, Memmott J: Studying insect diversity in the tropics. Philos Trans R Soc Lond B Biol Sci 1999;354:1811-1824.

16 Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PD: Wedding biodiversity inventory of a large and complex lepidoptera fauna with DNA barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1835-1845.

17 Valentini A, Pompanon F, Taberlet P: DNA barcoding for ecologists. Trends Ecol Evol 2009;24:110-117.

18 Blaxter ML: The promise of a DNA taxonomy. Philos Trans R Soc Lond B Biol Sci 2004;359:669-679.

19 Markmann M, Tautz D: Reverse taxonomy: An approach towards determining the diversity of meiobenthic organisms based on ribosomal rna signature sequences. Philos Trans R Soc Lond B Biol Sci 2005;360:1917-1924.

20 Mason B: Ecologists attack endangered-species logjam. Nature 2003;426:592.

21 Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD: DNA barcoding australia's fish species. Philos Trans R Soc Lond B Biol Sci 2005;360:1847-1857.

22 Ficetola GF, Miaud C, Pompanon F, Taberlet P: Species detection using environmental DNA from water samples. Biol Lett 2008;4:423-425.

23 Monaghan MT, Balke M, Gregory TR, Vogler AP: DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philos Trans R Soc Lond B Biol Sci 2005;360:1925-1933.

24 Smith MA, Fisher BL, Hebert PD: DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of madagascar. Philos Trans R Soc Lond B Biol Sci 2005;360:1825-1834.

25 Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PD: DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (diptera, tachinidae) are not all generalists. Proc Natl Acad Sci U S A 2007;104:4967-4972.

26 Wells JD: Development of DNA-based identification techniques for forensic entomology; phase 2, final report; in Justice USDo (ed. University of Alabama at Birmingham, 2002,

27 Pfenninger M, Nowak C, Kley C, Steinke D, Streit B: Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic chironomus (diptera) species. Mol Ecol 2007;16:1957-1968.

28 Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PD: DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (diptera: Tachinidae). Proc Natl Acad Sci U S A 2006;103:3657-3662.

29 Domingo-Roura Xea: Badger hair in shaving brushes comes from protected eurasian badgers. Biol Conserv 2006;128:6.

30 Lee JCI, Tsai LC, Yang CY, Liu CL, Huang LH, Linacre A, Hsieh HM: DNA profiling of shahtoosh. Electrophoresis 2006;27:3359-3362.

31 Teletchea F: Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. J ApplEcol 2008;45:967-975.

32 Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, McLaren P, Rodrigo AG, Lavery S, Baker CS: DNA surveillance: Web-based molecular identification of whales, dolphins, and porpoises. J Hered 2003;94:111-114.

33 Besansky NJ, Severson DW, Ferdig MT: DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. Trends Parasitol 2003;19:545-546.

34 Armstrong KF, Ball SL: DNA barcodes for biosecurity: Invasive species identification. Philos Trans R Soc Lond B Biol Sci 2005;360:1813-1823.

35 Nanney DL, Park C, Preparata R, Simon EM: Comparison of sequence differences in a variable 23s rrna domain among sets of cryptic species of ciliated protozoa. J Eukaryot Microbiol 1998;45:91-100.

36 Pace NR: A molecular view of microbial diversity and the biosphere. Science 1997;276:734-740.

37 Avise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA: Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, geomys pinetis. Proc Natl Acad Sci U S A 1979;76:6694-6698.

38 Avise JC, Lansman RA, Shade RO: The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus peromyscus. Genetics 1979;92:279-295.

39 Brown WM, Wright JW: Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus cnemidophorus). Science 1979;203:1247-1249.

40 de Graaf RM, van Alen TA, Dutilh BE, Kuiper JWP, van Zoggel HJAA, Huynh MB, Görtz H-D, Huynen MA, H.P.Hackstein. J: The mitochondrial genomes of the ciliates euplotes minuta and euplotes crassus. BMC Genomics 2009;10:514-541.

41 Pesole G, Gissi C, De Chirico A, Saccone C: Nucleotide substitution rate of mammalian mitochondrial genomes. J Mol Evol 1999;48:427-434.

42 Knowlton N, Weigt LA: New dates and new rates for divergence across the isthmus of panama. Proc R SocLond 1998;B 265:2257-2262.

43 Doyle JJ, Gaut BS: Evolution of genes and taxa: A primer. Plant Mol Biol 2000;42:1-23.

44 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R: DNA primers for amplification of mitochondrial cytochrome c oxidase subunit i from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 1994;3:294-299.

45 Zhang DX, Hewitt GM: Assessment of the universality and utility of a set of conserved mitochondrial coi primers in insects. Insect Mol Biol 1997;6:143-150.

46 Cox AJ, Hebert PD: Colonization, extinction, and phylogeographic patterning in a freshwater crustacean. Mol Ecol 2001;10:371-386.

47 Wares JP, Cunningham CW: Phylogeography and historical ecology of the north atlantic intertidal. Evolution 2001;55:2455-2469.

48 Simmons RB, Weller SJ: Utility and evolution of cytochrome b in insects. Mol Phylogenet Evol 2001;20:196-210.

49 Saitou N, Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406-425.

50 Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980;16:111-120.

51 Moritz C, Cicero C: DNA barcoding: Promise and pitfalls. PLoS Biol 2004;2:e354.

52 Lipscomb D, Platnick N, Wheeler QD: The intellectual content of taxonomy: A comment on DNA taxonomy. TRENDS in Ecology and Evolution 2003;18:65-66.

53 Mallet J, K. W: Taxonomy: Renaissance or tower of babel? TRENDS in Ecology and Evolution 2003;18:57-60.

54 Will KW, Rubinoff D: Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 2004;20:47-55.

55 Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM: Shortcuts in systematics?A commentary on DNA-based taxonomy. TRENDS in Ecology and Evolution 2003;18:63-66.

56 Lee MSY: The molecularisation of taxonomy Invertebrate Systematics 2004;18:1-7.

57 Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD: DNA barcodes distinguish species of tropical lepidoptera. Proc Natl Acad Sci U S A 2006;103:968-971.

58 Rubinoff D, Holland BS: Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst Biol 2005;54:952-961.

59 Ballard JW, Whitlock MC: The incomplete natural history of mitochondria. Mol Ecol 2004;13:729-744.

60 Bensasson D, Zhang D, Hartl DL, Hewitt GM: Mitochondrial pseudogenes: Evolution's misplaced witnesses. Trends Ecol Evol 2001;16:314-321.

61 Johnson NK, Cicero C: New mitochondrial DNA data affirm the importance of pleistocene speciation in north american birds. Evolution 2004;58:1122-1130.

62 Piganeau G, Gardner M, Eyre-Walker A: A broad survey of recombination in animal mitochondria. Mol Biol Evol 2004;21:2319-2325.

63 Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E: Widespread recombination in published animal mtdna sequences. Mol Biol Evol 2005;22:925-933.

64 Tatarenkov A, Avise JC: Rapid concerted evolution in animal mitochondrial DNA. Proc Biol Sci 2007;274:1795-1798.

65 Gibson TC, Kubisch HM, Brenner CA: Mitochondrial DNA deletions in rhesus macaque oocytes and embryos. Mol Hum Reprod 2005;11:785-789.

66 Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K: Mutation hot spots in mammalian mitochondrial DNA. Genome Res 2006;16:215-222.

67 Roe AD, Sperling FA: Patterns of evolution of mitochondrial cytochrome c oxidase i and ii DNA and implications for DNA barcoding. Mol Phylogenet Evol 2007;44:325-345.

68 Petri B, von Haeseler A, Paabo S: Extreme sequence heteroplasmy in bat mitochondrial DNA. Biol Chem 1996;377:661-667.

69 Farge G, Touraille S, Le Goff S, Petit N, Renoux M, Morel F, Alziari S: The nuclear genome is involved in heteroplasmy control in a mitochondrial mutant strain of drosophila subobscura. Eur J Biochem 2002;269:998-905.

70 Frey JE, Frey B: Origin of intra-individual variation in pcr-amplified mitochondrial cytochrome oxidase i of thrips tabaci (thysanoptera: Thripidae): Mitochondrial heteroplasmy or nuclear integration? Hereditas 2004;140:92-98.

71 Hilsdorf AW, Krieger JE: Restriction site heteroplasmy in the mitochondrial DNA of brycon opalinus (cuvier, 1819) (characiformes, characidae, bryconiae). Braz J Med Biol Res 2004;37:307-310.

72 Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL: Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish platichthys flesus. Mol Biol Evol 2002;19:2261-2264.

73 Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F: The complete mitochondrial DNA sequence of the basal hexapod tetrodontophora bielanensis: Evidence for heteroplasmy and trna translocations. Mol Biol Evol 2001;18:1293-1304.

74 Tsang WY, Lemire BD: Stable heteroplasmy but differential inheritance of a large mitochondrial DNA deletion in nematodes. Biochem Cell Biol 2002;80:645-654.

75 Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ: Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 1994;39:174-190.

76 Russo CA, Takezaki N, Nei M: Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol Biol Evol 1996;13:525-536.

77 Zardoya R, Meyer A: Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol 1996;13:933-942.

78 Erpenbeck D, Hooper JNA, Wörheide G: Co1 phylogenies in diploblasts and the 'barcoding of life' —are we sequencing a suboptimal partition? Molecular Ecology Notes 2005

79 Hebert PD, Gregory TR: The promise of DNA barcoding for taxonomy. Syst Biol 2005;54:852-859.

80 Bittner L, Halary S, Payri C, Cruaud C, de Reviers B, Lopez P, Bapteste E: Some considerations for analyzing biodiversity using integrative metagenomics and gene networks. Biol Direct 2010;5:5-47.

81 Rubinoff D, Cameron S, Will K: A genomic perspective on the shortcomings of mitochondrial DNA for "Barcoding" Identification. J Hered 2006;97:581-594.

82 Gatesy J, Milinkovitch M, Waddell V, Stanhope M: Stability of cladistic relationships between cetacea and higher-level artiodactyl taxa. Syst Biol 1999;48:6-20.

83 Baker RH, Gatesy J: Is morphology still relevant? EXS 2002:163-174.

84 Baker RH, DeSalle R: Multiple sources of character information and the phylogeny of hawaiian drosophilids. Syst Biol 1997;46:654-673.

85 Srivathsan A, Meier R: On the inappropriate use of kimura-2-parameter (k2p) divergences in the DNA-barcoding literature. Cladistics 2011;27:1-5.

86 Holder M, Lewis PO: Phylogeny estimation: Traditional and bayesian approaches. Nat Rev Genet 2003;4:275-284.

87 DeSalle R, Egan MG, Siddall M: The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1905-1916.

88 Elias M, Hill RI, Willmott KR, Dasmahapatra KK, Brower AV, Mallet J, Jiggins CD: Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc Biol Sci 2007;274:2881-2889.

89 Meyer CP, Paulay G: DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 2005;3:e422-434.

90 Kolisko M, Silberman JD, Cepicka I, Yubuki N, Takishita K, Yabuki A, Leander BS, Inouye I, Inagaki Y, Roger AJ, Simpson AG: A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats. Environ Microbiol 2010;12:2700-2710.

91 Kosakyan A, Heger TJ, Leander BS, Todorov M, Mitchell EA, Lara E: Coi barcoding of nebelid testate amoebae (amoebozoa: Arcellinida): Extensive cryptic diversity and redefinition of the hyalospheniidae schultze. Protist 2011:in press.

92 France SC, Hoover LL: DNA sequences of the mitochondrial coi gene have low levels of divergence among deep-sea octocorals (cnidaria: Anthozoa). Hydrobiologia 2002;471:149-155.

93 Shearer TL, Van Oppen MJ, Romano SL, Worheide G: Slow mitochondrial DNA sequence evolution in the anthozoa (cnidaria). Mol Ecol 2002;11:2475-2487.

94 Kerr KC, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PD: Comprehensive DNA barcode coverage of north american birds. Mol Ecol Notes 2007;7:535-543.

95 Gomez A, Wright PJ, Lunt DH, Cancino JM, Carvalho GR, Hughes RN: Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. Proc Biol Sci 2007;274:199-207.

96 Lefebure T, Douady CJ, Gouy M, Gibert J: Relationship between morphological taxonomy and molecular divergence within crustacea: Proposal of a molecular threshold to help species delimitation. Mol Phylogenet Evol 2006;40:435-447.

97 Remigio EA, Hebert PD: Testing the utility of partial coi sequences for phylogenetic estimates of gastropod relationships. Mol Phylogenet Evol 2003;29:641-647.

98 Spelda J, Reip HS, Oliveira-Biener U, Melzer RR: Barcoding fauna bavarica: Myriapoda - a contribution to DNA sequence-based identifications of centipedes and millipedes (chilopoda, diplopoda). Zookeys 2011:123-139.

99 Siddall ME, Kvist S, Phillips AJ, Oceguera-Figueroa AF: DNA barcoding of parasitic nematodes: Is it kosher? J Parasitol

100 Nijman V, Aliabadian M: Performance of distance-based DNA barcoding in the molecular identification of primates. C R Biol 2010;333:11-16.

101 Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, Rehner SA: Barcoding generalist predators by polymerase chain reaction: Carabids and spiders. Mol Ecol 2005;14:3247-3266.

102 Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, Wolfe AD, dePamphilis CW: Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc Natl Acad Sci U S A 2000;97:13166-13171.

103 Robba L, Russell SJ, Barker GL, Brodie J: Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (rhodophyta). Am J Bot 2006;93:1101-1108.

104 Saunders GW: Applying DNA barcoding to red macroalgae: A preliminary appraisal holds promise for future applications. Philos Trans R Soc Lond B Biol Sci 2005;360:1879-1888.

105 Saunders GW: Routine DNA barcoding of canadian gracilariales (rhodophyta) reveals the invasive species gracilaria vermiculophylla in british columbia. Mol Ecol Resour 2009;9 Suppl s1:140-150.

106 Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V: Land plants and DNA barcodes: Short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci 2005;360:1889-1895.

107 Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH: Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 2005;102:8369-8374.

108 Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SC: Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 2008;3:e2802.

109 Summerbell RC, Levesque CA, Seifert KA, Bovers M, Fell JW, Diaz MR, Boekhout T, de Hoog GS, Stalpers J, Crous PW: Microcoding: The second step in DNA barcoding. Philos Trans R Soc Lond B Biol Sci 2005;360:1897-1903.

110 Seifert KA: Progress towards DNA barcoding of fungi. Mol Ecol Resour 2009;9 Suppl s1:83-89.

111 Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Koljalg U: Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. PLoS One 2006;1:e59.

112 Nassonova E, Smirnov A, Fahrni J, Pawlowski J: Barcoding amoebae: Comparison of ssu, its and coi genes as tools for molecular identification of naked lobose amoebae. Protist 2011;161:102-115.

113 Heger TJ, Pawlowski J, Lara E, Leander BS, Todorov M, Golemansky V, Mitchell EA: Comparing potential coi and ssu rdna barcodes for assessing the diversity and phylogenetic relationships of cyphoderiid testate amoebae (rhizaria: Euglyphida). Protist 2010;162:131-141.

114 Scicluna SM, Tawari B, Clark CG: DNA barcoding of blastocystis. Protist 2006;157:77-85.

115 Evans KM, Chepurnov VA, Sluiman HJ, Thomas SJ, Spears BM, Mann DG: Highly differentiated populations of the freshwater diatom sellaphora capitata suggest limited dispersal and opportunities for allopatric speciation. Protist 2009;160:386-396.

116 MacGillivary ML, Kaczmarska I: Survey of the efficacy of a short fragment of the rbcl gene as a supplemental DNA barcode for diatoms. J Eukaryot Microbiol;58:529-536.

117 Moniz MB, Kaczmarska I: Barcoding of diatoms: Nuclear encoded its revisited. Protist;161:7-34.

118 Moniz MB, Kaczmarska I: Barcoding diatoms: Is there a good marker? Mol Ecol Resour 2009;9 Suppl s1:65-74.

119 Hamsher SE, Evans KM, Mann DG, Poulickova A, Saunders GW: Barcoding diatoms: Exploring alternatives to coi-5p. Protist 2011;162:405-422.

120 Lin S, Zhang H, Hou Y, Zhuang Y, Miranda L: High-level diversity of dinoflagellates in the natural environment, revealed by assessment of mitochondrial cox1 and cob genes for dinoflagellate DNA barcoding. Appl Environ Microbiol 2009;75:1279-1290.

121 Murray S, Ip C, Moore R, Nagahama Y, Fukuyo Y: Are prorocentroid dinoflagellates monophyletic?A studyof 25 species based on nuclear and mitochondrial genes. Protist 2009;160:245-255.

122 Stern RF, Horak A, Andrew R, Coffroth MA, Andersen RA, Küpper FC, Jameson I, Hoppenrath M, Véron B, Kasai F, Brand J, James ER, Keeling PJ: Environmental barcoding reveals massive dinoflagellate diversity in marine environments. PLoS One 2010;5

123 Barth D, Krenek S, Fokin SI, Berendonk TU: Intraspecific genetic variation in paramecium revealed by mitochondrial cytochrome c oxidase i sequences. J Eukaryot Microbiol 2006;53:20-25.

124 Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N, Ikonomi P: Barcoding ciliates: A comprehensive study of 75 isolates of the genus tetrahymena. Int J Syst Evol Microbiol 2007;57:2412-2425.

125 Kher CP, Doerder FP, Cooper J, Ikonomi P, Achilles-Day U, Kupper FC, Lynn DH: Barcoding tetrahymena: Discriminating species and identifying unknowns using the cytochrome c oxidase subunit i (cox-1) barcode. Protist;162:2-13.

126 Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N, Ikonomi P: Barcoding ciliates: A comprehensive study of 75 isolates of the genus tetrahymena. Int J Syst Evol Microbiol 2007;57:2412-2425.

127 Corliss JO: Biodiversity and biocomplexity of the protists and an overview of their significant roles in maintenance of our biosphere. Acta Protozool 2002;41:199-119.

128 Moreira D, Lopez-Garcia P: The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol 2002;10:31-38.

129 Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF: The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 2005;52:399-351.

130 Stoeck T, Taylor GT, Epstein SS: Novel eukaryotes from the permanently anoxic cariaco basin (caribbean sea). Appl Environ Microbiol 2003;69:5656-5663.

131 Bass D, Cavalier-Smith T: Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of cercozoa (protozoa). Int J Syst Evol Microbiol 2004;54:2393-2404.

132 Countway PD, Gast RJ, Savai P, Caron DA: Protistan diversity estimates based on 18s rdna from seawater incubations in the western north atlantic. J Eukaryot Microbiol 2005;52:95-106.

133 Caron DA: Past president's address: Protistan biogeography: Why all the fuss? J Eukaryot Microbiol 2009;56:105-112.

134 Struder-Kypke MC, Wright AD, Foissner W, Chatzinotas A, Lynn DH: Molecular phylogeny of litostome ciliates (ciliophora, litostomatea) with emphasis on free-living haptorian genera. Protist 2006;157:261-278.

135 Finlay BJ, Fenchel T: Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 2004;155:237-244.

136 Boenigk J, Pfandl K, Garstecki T, Harms H, Novarino G, Chatzinotas A: Evidence for geographic isolation and signs of endemism within a protistan morphospecies. Appl Environ Microbiol 2006;72:5159-5164.

137 Leander BS, Kuvardina ON, Aleshin VV, Mylnikov AP, Keeling PJ: Molecular phylogeny and surface morphology of colpodella edax (alveolata): Insights into the phagotrophic ancestry of apicomplexans. J Eukaryot Microbiol 2003;50:334-340.

138 Slapeta J, Lopez-Garcia P, Moreira D: Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol Biol Evol 2006;23:23-29.

139 Adl SM: Diversity, nomenclature, and taxonomy of protists. Sysl Biol 2007;56:684-689.

140 Dini F, D. N: Growth rates of marine ciliates on diverse organisms reveal ecological specializations within morphospecies. Microb Ecol 1999;37:13-23.

141 Petroni G, Dini F, Verni F, Rosati G: A molecular approach to the tangled intrageneric relationships underlying phylogeny in euplotes (ciliophora, spirotrichea). Mol Phylogenet Evol 2002;22:118-130.

142 Di Giuseppe G, Erra F, Dini F, Alimenti C, Vallesi A, Pedrini B, Wuthrich K, Luporini P: Antarctic and arctic populations of the ciliate euplotes nobilii show common pheromone-mediated cell-cell signaling and cross-mating. Proc Natl Acad Sci U S A 2010;108:3181-3186.

143 Katoh K, Kuma K, Miyata T, Toh H: Improvement in the accuracy of multiple sequence alignment program mafft. Genome Inform 2005;16:22-33.

144 Katoh K, Misawa K, Kuma K, Miyata T: Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res 2002;30:3059-3066.

145 Abascal F, Zardoya R, Telford MJ: Translatorx: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res 2010;38:W7-13.

146 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: Mega5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731-2739.

147 Schmidt H, Strimmer K, Vingron M, von Haeseler A: Tree-puzzle: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 2002, 18:502-504 2002;18:502-505.

148 Strimmer K, von Haeseler A: Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci USA 1997;94:6815-6820.

149 Negrisolo E, Kuhl H, Forcato C, Vitulo N, Reinhardt R, Patarnello T, Bargelloni L: Different phylogenomic approaches to resolve the evolutionary relationships among model fish species. Mol Biol Evol 2010;27:2757-2774.

150 Huelsenbeck JP, Ronquist F: Mrbayes: Bayesian inference of phylogeny. Bioinformatics 2001;17:754-756.

151 Stamatakis A: Raxml-vi-hpc: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688-2690.

152 Silvestro D, Michalak I: Raxmlgui: A graphical front-end for raxml. Organisms diversity and evolution. 2011

153 Yang Z (ed Computational molecular evolution, 2006, pp 376.

154 Felsenstein J: Confidence limits on phylogenies: An approach using bootstrap. Evolution 1985;39:783-792.

155 Thompson JD, Higgins DG, Gibson TJ: Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673-4680.

156 Vallesi A, Giuseppe GD, Dini F, Luporini P: Pheromone evolution in the protozoan ciliate, euplotes: The ability to synthesize diffusible forms is ancestral and secondarily lost. Mol Phylogenet Evol 2008;47:439-442.

157 Yi Z, Song W, Clamp JC, Chen Z, Gao S, Zhang Q: Reconsideration of systematic relationships within the order euplotida (protista, ciliophora) using new sequences of the gene coding for small-subunit rrna and testing the use of combined data sets to construct phylogenies of the diophrys-complex. Molecular Phylogenetics and Evolution 2009;50:599-507.

Sarzana, 29 Febbraio 2012

Il dottorato è temporalmente, psicologicamente, mentalmente, fisicamente, subconsciamente, misticamente, ed onicamente un periodo di tempo lungo e significativo, ne consegue che le persone da ringraziare sono davvero tante.

La prima persona che ringrazio, non per dovere ma per sentita gratitudine, è il Dr. Ph.D. Graziano Di Giuseppe. Errare è umano, ma perseverare è diabolico. Come ha detto qualcuno, io sono stata ributtata dentro con "un colpo di tacco", ed il secondo tentative è andato bene nella stessa misura in cui era andato male il primo. Ora non ho né la lucidità né la lungimiranza per asserire se per il mio futuro fare il dottorato sia stato un bene o un male, magari passerò la vita ad invidiare quelle che dopo la terza media sono andate a fare le estetiste. Indipendentemente da questo, se dottorato doveva essere, quello su cui non ho il benché minimo dubbio, è che a farlo sotto la supervisione di una persona di alto valore sia umano che scientifico quale è il Dr. Ph.D. Graziano Di Giuseppe è stato un privilegio.

Non me ne voglia il Professor Dini se lo inserisco come secondo nome. Trattare con un Professore con un secolo di esperienza, ma contemporaneamente con la mente di un giovane, ha reso molte cose più facili e gestibili. Ogni volta che ho avuto bisogno di lui, lui c'é stato. Senza i suoi interventi di altissimo livello culturale il laboratorio sarebbe sprofondato nella monotonia. Farò Tesoro prezioso di tutti i suoi insegnamenti, sia di quelli scientifici ma soprattutto di quelli di vita. Quando mi si presenteranno davanti situazioni difficili, ho la soluzione imbottigliata a portata di mano. Una promessa, però, è una promessa. La aspetto al rinfresco presso la "Locanda della Bacchettona", e questa volta MI DEVE chiamare Dottoressa.

Il terzo ringraziamento è per la persona con la quale ho avuto meno a che fare di tutte. Non mi ricordo neppure che faccia abbia, l'ho visto una volta sola. Tuttavia, TEMO di aver sentito più volte al telefono lui negli ultimi 3 mesi che mia madre in 6 mesi di Erasmus. Sto parlando del Dr.Ph.D. Enrico Negrisolo, ricercatore presso l'Università di Padova. Enrico mi ha assistita (sì, assitere è il verbo giusto) per un periodo di tempo molto limitato, ma il più significativo, quello dell'analisi dei dati. Confesso che quasi tutte le volte, dopo aver chiuso skype, avrei voluto essere a Parigi per buttarmi per rotolare giù dalla scalinata di Mont Martre. Tralasciando i miei "brain storming", sono stata proprio fortunata a trovare una persona così competente ma altrettanto alla mano, sulla quale ho potuto sempre contare, anche per le banalità (banalità per lui, brain storming per me).

Ringrazio il Dr.Ph.D. Fabrizio Erra ed il Sig. Francesco Frontini per la loro competenza e perché sono sempre stati veramente molto disponibili nei miei confronti. E grazie anche per tutti i caffé che gli ho scroccato al Polo Nobili....e già che ci siamo grazie anche al caffé!!!

Ringrazio la simpatica allegra combriccola della SIP onlus, scienziati di altissimo livello ed umanamente persone squisite. Con loro mi sono proprio divertita ai vari congressi (a cui ho partecipato grazie al mio tutor che me li ha finaziati interamente, altro problema del dottorato italiano, oltre a quello del tutoraggio).

Ringrazio il personale strutturato e non dell'Unità di Protistologia dell'Università di Pisa per la loro compagnia e cortesia.

Ringrazio le tesiste che mi hanno fatto compagnia durante questo percorso per i loro impegno, serietà e simpatia: Chiara, Céline, Laura P., Caterina, Laura A. e Giulia.

Ringrazio i miei primi mentori, i Dr.Ph.D. Davide Cervia e Chiara Ristori. Grande Chiara!Io e te ci siamo divertite tanto insieme, mi sei proprio rimasta nel cuore!!!!Ringrazio anche i loro colleghi i Dr.Ph.D. Davide Martini, Maurizio Cammalleri, Massimo Dal Monte e tutto il personale non strutturato di "quel posto".

Ringrazio le mie compagne di (s)ventura in ordine alfabetico Caterina, Lisa e Silvia per il supporto morale.

Non mi è mai piaciuto, così, per motivi di riservatezza, scrivere i ringraziamenti personali, infatti nelle tesi delle lauree triennali e specialistica non l'avevo fatto, ma arrivare in fondo è stata davvero dura, così dura che non posso proprio definirlo come un traguardo personale, bensì collettivo.

Mi ritengo una persona fortunata perché ho una bellssima famiglia, due genitori fantastici che vicini o lontanti, mi sostengono sempre. E sopportata. E ce ne vuole. Ringrazio anche il nuovo capo famiglia, la cagnolina più dolce ed affettuosa del mondo. Il giorno in cui l'ho trovata al canile ho vinto la lotteria.

Ringrazio Massimo, per tutto. Per starmi sempre vicino, per mettermi sempre al primo posto, per aver sempre dato importanza al mio lavoro, per non aver mai formalmente preteso che facessi tutti io in casa perché sono a casa, per non avermi minimamente fatto pesare tutti i week-end in casa e le serate a sbattersi la testa davanti al PC. Soprattutto, grazie per avermi sopportata. E ce ne vuole veramente tanta di pazienza per sopportarmi!

Ringrazio la mia amica storica Milena. D'altronde se non inserivo nei ringraziamenti Milena neppure questa volta, era la volta buona che mi avrebbe mandata a quel paese per sempre. Ed io ti ringrazio proprio per questo, perché mi mandi a quel paese quando serve. Considero la nostra amicizia come uno dei più grandi successi della mia vita.

Ringrazio il mio amico storico Davide, che mi è sempre stato vicino.

Ringrazio tutte i cari amici che, con modi diversi, una e-mail, un messaggino su whatsapp, su facebook ecc... mi hanno sempre fatto sapere che mi pensano e mi hanno spronata a non mollare. I nomi (oddio, ero abituata a fare le liste dei nomi di Euplotes!!!): Ilaria, Daria, Giada, Susanna, Graziella, Silvia, Leonardo e Serena.

Ringrazio per la compagnia gli amici della penultima carrozza del treno delle 07:57.

Ringrazio i simpatico amici della palestra Circuit Sarzana. L'ambiente di quella palestra mi piace molto, e l'ambiente lo fanno le persone che la frequentano.

Una menzione speciale per Nunzia per avermi fatto i conti della consegna,e avermi fatto notare che non avevo capito assolutamente niente :-P

Ringrazio i miei amori musicali Placebo e Muse, tutte le varie saghe vampiresche, e il nocciolato novi, che mi hanno tenuto compagnia.

E grazie anche a me stessa.

E alle persone che pensavano e/o speravano che non ce l'avrei fatta:

prrrrrrrrrrrrrrr